A pre-evacuation database for use in egress simulations

Lovreglio, R., Kuligowski, E., Gwynne, S., & Boyce, K. E. (2019). A pre-evacuation database for use in egress simulations. *Fire Safety Journal*, *105*, 107-128. https://doi.org/10.1016/j.firesaf.2018.12.009

Published in:
Fire Safety Journal

Publication Status:
Published (in print/issue): 01/04/2019

DOI:
10.1016/j.firesaf.2018.12.009

Document Version
Author Accepted version

A Pre-Evacuation Database for Use in Egress Simulations

Lovreglio R.[a], Kuligowski E.[b], Gwynne S.[c], and Boyce K.[d]

[a] Massey University, New Zealand
[b] National Institute of Standard and Technology, USA
[c] Movement Strategies, UK
[d] Ulster University, UK

Abstract:
Quantifying the pre-evacuation time (i.e., the time between first awareness and deliberate evacuation movement), is a key task for evacuation model users and fire safety engineers. The identification and employment of pre-evacuation data given an incident scenario is not a simple task for evacuation model users and fire safety engineers since data is typically scarce, partial and often difficult to access. In this work, we address this issue by presenting an expanded database including pre-evacuation times collected from 9 fire incidents and 103 evacuation drills involving 13,591 evacuees in 16 countries. These case studies are grouped according to the occupancy type of the structure(s) involved. We also used cluster analysis to identify sub-groups and potential factors that influence performance. Using this pre-evacuation data, we calibrate a set of pre-evacuation distributions that can be used to represent pre-evacuation data in existing building evacuation models. This work provides a useful resource for evacuation model users and fire safety engineers and also may provide additional insights to researchers into the factors that influence pre-evacuation times. Finally, this work can have an impact on future data collection and analysis by identifying the need for new data for specific occupancies.

Keywords: building evacuation, pre-evacuation, database, egress modelling

1. INTRODUCTION

Evacuee behavior is a key factor in any fire safety performance based design. As such, understanding and predicting evacuee behavior is fundamental to enhancing the safety of buildings. To date, several egress models have been developed to simulate fire evacuations to determine whether the required safe egress time is less than the available safe egress time [1,2].

The building evacuation process can be divided into several components that form a response timeline [3,4]. The total evacuation time is generally divided into pre-evacuation time and travel time. The *pre-evacuation time* is the interval between the time at which a general alarm signal or warning is given (or other cues received) and the time at which the first deliberate evacuation movement is made [5]. The *travel time* is the interval needed for an evacuee to reach a safe place, once movement toward an exit has begun [5]. The evacuation performance in these two stages is dependent on the conditions faced by evacuees and their capacity to respond. As such, evacuation model users and fire safety engineers need to identify those conditions that define the incident scenarios that they wish to simulate [6].

A key issue for model users is the identification and employment of pre-evacuation data that describes likely egress performance given pre-determined fire scenarios. This task might be demanding as pre-evacuation data is typically scarce, partial, difficult to access and/or in a format which is not supported by evacuation models [6]. An initial attempt to simplify this process was made by Gwynne and Boyce in the

SFPE Handbook [4] who pulled together existing engineering data in a series of tables in a format which could be easily accessed by evacuation modelers. Part of the work done by Gwynne and Boyce focused on pre-evacuation data. They identified 76 case studies (i.e., 4 fire incidents and 72 evacuation drills) which they divided according to occupancy classes and presented in nine tables. The goal of the SFPE chapter is to provide the user with a characterization of each case study and present the related pre-evacuation data using four descriptive statistics: mean, standard deviation, minimum and maximum (depending on the data summary available in the original source material). The first limitation of this existing database is that, in many instances, one or more of the four statistics is missing. Moreover, the statistics presented in the existing database are not intended to be used directly as an input to existing evacuation models, which require users to define the parameters of pre-evacuation distributions. These summary measures were instead intended to encourage the reader to select between and explore the original data sources provided.

In this paper, we address this issue presenting an expanded version of the database proposed by Gwynne and Boyce [4]. We present an expanded database, including pre-evacuation times collected from 9 fire incidents and 103 evacuation drills[1]. In contrast to the work presented in the SFPE chapter, we collected raw pre-evacuation data such as individual pre-evacuation data or aggregated pre-evacuation data through cumulative frequencies from the original datasets. This was achieved by searching the original references and contacting the original researchers. It was possible to collect data from the original references for 91 case studies while the data for the remaining case studies were provided by the authors. Through this process, we collected 2889 data points, where each data point includes a pre-evacuation time and its frequency. The advantage of this new data structure, i.e., combining pre-evacuation times and frequencies is that it allows (1) the calculation of all the statistics proposed in this paper without missing values; and (2) the calibration of pre-evacuation distributions for single case studies or a combination of them.

The case studies included in this expanded database are categorized according to the occupancy involved (e.g., business, residential, mercantile, etc.), in accordance with the work conducted by Gwynne and Boyce in the SFPE Handbook [4]. This categorization was originally used to reflect the typical way in which practitioners would search for and select data; i.e., the first decision being the type of occupancy being represented. The original occupancy classification has been expanded to accommodate the new case studies identified in this work i.e., Hotel, Road Tunnel (i.e., drivers' evacuation behavior) and Miscellaneous. This was done to accommodate the new case studies identified in this work. A few original occupancy classes such as Industrial, Health Care and Transport (i.e., transportation terminal users) are not included in this work as they included only a few case studies that are included in the new Miscellaneous class. Here, we go further than the original approach by examining variables that further differentiate the data; i.e., we use cluster analysis to identify possible sub-sets of data within the occupancy groups in an attempt to explore additional influential factors within each occupancy type. Such an analysis allows the identification of possible factors that may influence pre-evacuation timing with a view to potentially producing more rational approaches to grouping data in the future (i.e., beyond occupancy grouping).

[1] An evacuation drill is defined as a preplanned simulation of an emergency evacuation for a specific incident scenario [26]. In this work, evacuation drills refer to both unannounced and announced evacuations

2. BACKGROUND

To date, numerous studies have shown that pre-evacuation time can represent a significant portion of an evacuee's total evacuation time [7–9], which can have serious consequences depending on the nature of the incident [8]. Conceptual models have been developed to show the types, and even the sequence, of evacuee behaviors that are performed during pre-evacuation. One of the first of these models was proposed by Canter, Breaux, and Sime [10] who charted processes for all the possible actions and responses that could be taken by evacuees in different types of occupancies [11]. More recent conceptual models have been proposed in [12–14] which were inspired by the general Protective Action Model by Lindell and Perry [15]. Other studies have been performed to investigate and quantify the internal and external factors that might influence pre-evacuation behavior [11,16,17]. Notwithstanding, pre-evacuation behaviors are generally less documented, quantified and modelled than movement behaviors [8,18,19].

At a conceptual level, three main modelling approaches have been proposed to represent evacuee pre-evacuation time in building evacuation models [13]. The first approach relies on the user assignment of a pre-defined time to individuals or groups (i.e., a deterministic approach) or a pseudo-random number obtained from a distribution (i.e., a stochastic approach). The second approach involves the user assignment of sequences of pre-evacuation actions. The agents move to different parts of the simulated building to perform their activities. Each action has a pre-defined specific duration for each agent, assigned by the user. The last approach is the predictive-based approach. In this case, agents perform protective actions in accordance with different internal and external factors. Examples of this last approach are the Evacuation Decision Model proposed in [20] and its implementations for different evacuation studies [21–24].

All three approaches have strengths and limitations. For instance, the main weakness of the first two approaches is that the behavior is not actually predicted by the models, but it is based on user assumptions while the main limitation of the third approach is the 'homogeneity' assumption (i.e., agents react to particular cues in similar ways) [13]. The main advantage of the predictive-based approach (i.e., the third approach) is that the evacuee pre-evacuation behavior is actually modelled whilst the other two approaches expect users to define such behavior as an input by selecting a pre-evacuation distribution(s) or a sequence of pre-evacuation actions. It is recognized, however, that such an approach has only been applied to a limited number of cases and situations such as [21–24]. From an implementation viewpoint and evacuation model users still typically rely on the first approach to simulate pre-evacuation timing (particularly the stochastic approach).

The first approach (either deterministic or stochastic) requires the model users to supply pre-evacuation timing data to implicitly represent the types of behaviors that various people perform during the pre-evacuation time period, and the overall time that it takes to perform these series of behaviors. This is done by asking the users to define pre-evacuation distributions which could represent the timing uncertainty. Data from evacuation drills and real emergencies can be used to quantify the pre-evacuation time and distributions for different types of buildings [25,26]. To date, three pre-evacuation databases have been proposed to do just that. The first, produced by Shi et al. [27], introduced several descriptive statistics for pre-evacuation times collected in 69 evacuation drills dividing them according to a small set of occupancy classes. The second database produced by Fahy and Proulx [28], provided users with descriptive statistics for several fire drills and fire accidents which had occurred in offices, residential buildings, hotels and stores. The most recent database was provided by Gwynne and Boyce in the SFPE Handbook [4]. They identified 76 case studies and provided descriptive statistics for pre-evacuation time divided according to occupancy classes. The main similarity and limitation of the existing works is that all provide only descriptive statistics of pre-evacuation data. Although these statistics provide users with the order of magnitude of the pre-evacuation times for different scenarios, converting them into a probabilistic distribution is not an easy task for evacuation model users. Ideally, evacuation model users should find those distributions in the original references listed in the existing databases. However, many times those are not published as they are out of the scope of those works. In this article, we aim to fill this gap expanding the pre-evacuation database proposed by Gwynne and Boyce in the SFPE Handbook [4] and estimating those distributions using the methodological approach described in the following section.

3. METHODOLOGY

3.1 Data Selection and Representation

The pre-evacuation data included in this paper has been found in sources typically considered as credible outlets within the field. Those sources were identified by Gwynne and Boyce in the SFPE Handbook [4], and include:

1. Journal publications: Journal of Fire Protection Engineering, Fire Safety Journal, Fire Technology, Fire and Materials, Safety Science, International Journal of Performance- Based Fire Codes, Journal of Applied Fire Sciences, Building and Environment, Journal of Transportation Engineering Transportation Research Record, Physica A:
2. Conference proceedings: International Association Fire Safety Science (IAFSS), Interflam, Pedestrian and Evacuation Dynamics (PED), Human Behavior in Fire, Asia- Oceania Association for Fire and Technology, Mobility and Transport for Elderly and Disabled People;
3. Reports: National Institute of Standards and Technology (NIST), National Fire Protection Association (NFPA), National Research Council Canada (NRCC), British Standard Institute (BSI), Fire Protection Research Foundation (FPRF), Lund Department of Fire Safety Engineering and VTT Technical Research Centre (Finland).

The identification of potential references has also been assisted by the reviews provided in seminal publications in the field of human behavior in fire [4,27–29]. The final set of references which provided data for analysis here were identified and selected according to the following criteria, whereby the sources were:

1. Publicly available;
2. Written in English (or where translations were available upon request);
3. Published after 1980 to ensure relevance;
4. Providing at least four data points, where each data point included a pre-evacuation time and its cumulative frequency (i.e., percentiles), in digital or graphical forms.

The first three criteria are the same as those adopted by Gwynne and Boyce in the SFPE Handbook [4] to identify papers presenting pre-evacuation data. The reason behind the selection of data published after 1980 is to get evacuation data which is fairly contemporary and thus having evacuation conditions and responses similar to those present today. The last criterion was added in this work to select papers having a specific pre-evacuation data structure. This has been added since the goal here, in contrast to the SFPE

Chapter, is to collect raw pre-evacuation data (individual pre-evacuation data or aggregated pre-evacuation data through frequencies). This fourth criterion is fundamental to broadly estimate pre-evacuation distributions (as explained in Section 3.3), that are more useful for computational evacuation models. In fact, having four percentiles allows the estimation of distributions that represent the pre-evacuation trends observed during an evacuation.
The pre-evacuation data (i.e., pre-evacuation times and their frequencies) is generally available in the published references through tables and charts. When a reference did not satisfy the fourth criterion, the reference authors were contacted and asked if they could provide the raw data. If the data was available only in graphic form, it was converted into a digital form by using the open source application called WebPlotDigitizer[2]. This graphic conversion was done for 34 case studies which are identified by the * symbol in the following Tables 2,4,8,12, 14 and 16. The selection and rejection procedure is illustrated in Figure 1.

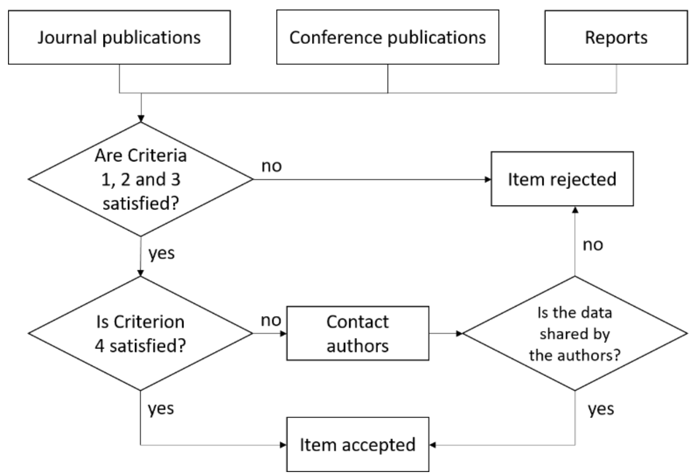

Figure 1 – Data selection procedure.

The datasets (originating from evacuation drills or fire incidents) included in the expanded database are presented in this paper in tabular and graphical form. The tables provide two statistics regarding the pre-evacuation times: mean and standard deviation. Those statistics were taken from the references when available, otherwise they were calculated using the frequencies either provided in the paper or by the reference authors.

The tables presented here also provide background information that allows the reader to understand the context in which the data was collected and the scenarios associated with the datasets.

In this paper, the tables are presented in standardized format according to the structure used in the SFPE Chapter [4]. The table structure includes the following items:

- Reference ("Ref." in the Tables) indicates one or more references where the data is from;

[2] Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

- Building is a description of the type of building where the evacuation took place;
- Country in which the building is located;
- Nature indicates the nature of the evacuation and the understanding of the event by the population; i.e., whether the evacuation was a total evacuation or a partial evacuation and whether it took place during announced or unannounced drills/real events (i.e., total Unannounced Drill: UD; Partial Unannounced Drill: P-UD), Announced Drills (AD), Semi-Announced Drills (SAD) or Fire Incidents (FI);
- Alarm provides information regarding the alarm systems: i.e., sirens, bells, and horns (AL); T3 fire alarm systems (T3); live voice notifications (LV); and prerecorded voice notifications (PV);
- Floors indicate the number of floors of the building;
- Sample is the number of evacuees whose pre-evacuation times are observed and included in the study;
- Mean and Standard Deviation (SD) of the pre-evacuation times[3];

The datasets included in this publication are also presented in a graphical form, i.e., in two standard charts: a 2D plot showing the pre-evacuation time on the horizontal axis and the cumulative frequency or probability on the vertical axis and a 2D plot showing the mean and standard deviation of pre-evacuation times in the horizontal and vertical axis respectively. The former chart form has been selected as cumulative measurements easily allow the comparison of several pre-evacuation datasets and distributions. Moreover, such a format is preferred as the pre-evacuation distributions are estimated using cumulative frequencies as explained in Section 3.2. The second form of chart is used to illustrate where several datasets are located in such a two-dimensional space. This approach enabled cluster analysis to be performed (as described in Section 3.3) to identify potential factors affecting the grouping.

3.2 Distribution Estimation

Starting from collected pre-evacuation data points, it is possible to calibrate continuous distributions. Let *f(x*$|a, b)$ be a continuous probabilistic distribution defined by two parameters (i.e., a and b) and *F(x*$|a, b)$ its cumulative distribution. Let (t_i^d, P_i^d) be a data point representing the *i* pre-evacuation times (t_i^d) and the cumulative frequencies (P_i^d) of the *d* dataset (*i=1, …,* I^d *and d=1, …, D*). *f* can be calibrated with the existing data using Least Squares Method [30] by solving the optimization problem in Equation 1.

$$\arg\min_{a,b} \sum_{d=1}^{D} w_d \sum_{i=1}^{I^d} (P_i - F(x_i|a, b))^2 \qquad \text{Equation 1}$$

[3] Those statistics were taken from the original references (when available) or calculated using the available frequency data:

$$mean = \frac{\sum_i t_i E_i}{\sum_i E_i}$$

$$st.dev. = \sqrt{\frac{\sum_i (t_i - mean)^2 E_i}{\sum_i E_i}}$$

where E_i is the number of evacuees having t_i as pre-evacuation time.

where w_d are the weights associated with each dataset. Considering that each *d* dataset is made of I^d data points accounting for e^d evacuees, we assume that w_d is the ratio between $e^d$and I^d (i.e., $w_d = e^d/I^d$). In such a way, each dataset contributes in the fitting of a curve according to the size of the sample. Further details regarding such an approach (i.e., weighted nonlinear regression) and its challenges are available in [31,32].

In this paper, we consider four possible pre-evacuation distributions defined by two parameters: gamma (Equation 2), lognormal (Equation 3), loglogistic (Equation 4), and Weibull distributions (Equation 5). These are fitted against the various dataset groupings. Those distribution were selected as they are defined only for positive values of the *x* random variable (i.e., negative value of pre-evacuation time are not allowed). Those distribution also have a skewed shape which is typical for pre-evacuation data [7]. Moreover, those distributions are implemented in many well-known evacuation models such as FDS+Evac [33], Pathfinder [34] and EXODUS [35].

$$\text{Gamma: } F(x|a,b) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{\frac{-t}{b}} dt \qquad \text{Equation 2}$$

$$\text{Lognormal: } F(x|a,b) = \frac{1}{b\sqrt{2\pi}} \int_0^x \frac{\exp(\frac{-(\ln(t)-a)^2}{2b^2})}{t} dt \qquad \text{Equation 3}$$

$$\text{Loglogistic: } F(x|a,b) = \frac{1}{1+(x/a)^{-b}} \qquad \text{Equation 4}$$

$$\text{Weibull: } F(x|a,b) = \int_0^x b a^{-b} t^{b-1} \exp(-(t/a)^b)\ dt \qquad \text{Equation 5}$$

The fitting of those four distributions is assessed and compared using the R^2 parameter[4]. The results of all four distributions are always presented (in Section 4) as some evacuation models might not have the capacity to represent each distribution. As such, the users can select an alternative distribution between the remaining ones.

The proposed approach can be used to estimate more complex distribution having more than two parameters. However, those distributions have not been used in the literature or implemented in well-known evacuation models.

The methodology proposed in this section is used in Section 4 to calibrate pre-evacuation distributions by combining the data points from different data sets.

[4] There is some debate over the general applicability of the R^2 indicator, especially to non-linear distributions. Multiple fitting indicators can be used when Likelihood methods are used to fit models as indicated in [80]. However, in this work we adopted a weighted Least Squares method where the weights account for the sample size of each dataset (see Equation 1). As such, for the purpose of this paper, readers should refer to R^2 indicators as well as to the regression charts (i.e. the scatter plots including the lines representing regression models) to have a comprehensive understanding of advantages and limitations of the proposed regression models.

3.3 Clustering Analysis

In this work, a clustering solution was used to group the case studies identified in the expanded database. To achieve this, the k-mean cluster analysis was used [36]. Such a mathematical solution allows us to investigate whether it is possible to subdivide the case studies included in our database into clusters and thus identify candidate factors that may segregate the datasets in addition to or instead of the occupancy types already identified.

The average and standard deviation of the pre-evacuation times of each dataset belonging to the same occupancy group are the input for the cluster analysis. Let $\boldsymbol{x_i} = \{M_i, S_i\}$ be a two-dimensional real vector defined by the mean (M_i) and standard deviation (S_i) of the *i* case study (i=1, …, n). The k-mean cluster analysis is a technique that allows the partition of the *n* vectors into *k* (<n) cluster $\boldsymbol{C} = \{C_1, \dots, C_k\}$ by minimizing the within-cluster sum of squares [36]. This is done by solving this optimization problem:

$$\arg\min_{\boldsymbol{C}} \sum_{i=j}^{k} \sum_{x \in C_j} \|\boldsymbol{x} - \boldsymbol{\mu_j}\|^2 \qquad \text{Equation 2}$$

where $\boldsymbol{\mu_j}$ is the mean of the $\boldsymbol{x_i}$ points belonging to the C_j cluster. A fundamental input requirement of this clustering approach is the number of clusters (*k*). This number can be identified using the Elbow method [37], which focuses on the reduction of the within-cluster sum of square as a function of the number of clusters as illustrated in Figure 2. As such, the resulting clusters includes case studies sharing 'similar' mean and standard deviation of pre-evacuation time.

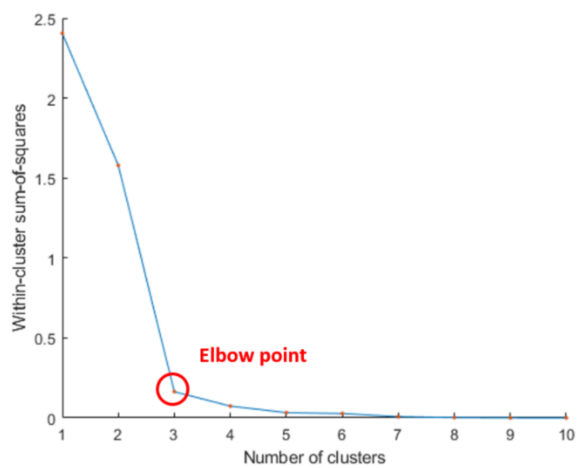

Figure 2 – An example of Elbow chart.

In the following sections, the pre-evacuation distributions in Section 3.2 are calibrated for each individual cluster. It is worth highlighting that the R^2 parameters given below must not be used as a criterion to select between clusters, but need to be only used to choose between distributions estimated within the same cluster.

4. RESULTS

The database presented in this paper is an expanded version of the database proposed by Gwynne and Boyce [4]. It includes pre-evacuation times collected from 112 case studies, including 9 fire incidents and 103 evacuation drills. 93 of those drills were unannounced while the remaining 10 drills were announced or semi announced. The data included in this dataset originates from 16 countries. The percentage of case studies belonging to each country is illustrated in Figure 3.a. It is evident that Sweden and the US account for almost 40% of the total case studies.

The case studies are divided here depending on their occupancy, in accordance with the original work conducted by Gwynne and Boyce in the SFPE Handbook [4]. The original occupancy classification has been partially extended in this work by the introduction of 3 new categories, i.e., Hotel, Road Tunnel and Miscellaneous. The list of occupancy groups and the percentage of case studies belonging to each group is depicted in Figure 3.b. From Figure 3.b it is evident that a third of the case studies consist of evacuations that took place in educational buildings.

A comparison of the expanded database with the databases proposed by Gwynne and Boyce [4], Fahy and Proulx [28] and Shi et al. [27] is presented in Figure 4[5].

From Figure 4, it is possible to observe some overlapping between the expanded database and the existing databases. 68% of case studies from the Gwynne and Boyce database [4] are included in the expanded database. Moreover, all the case studies from the Fahy and Proulx [28] and 9% from Shi et al. [27] are also included in the expanded database. The case studies that were excluded did not meet the criterion 4 (See Section 3.1).

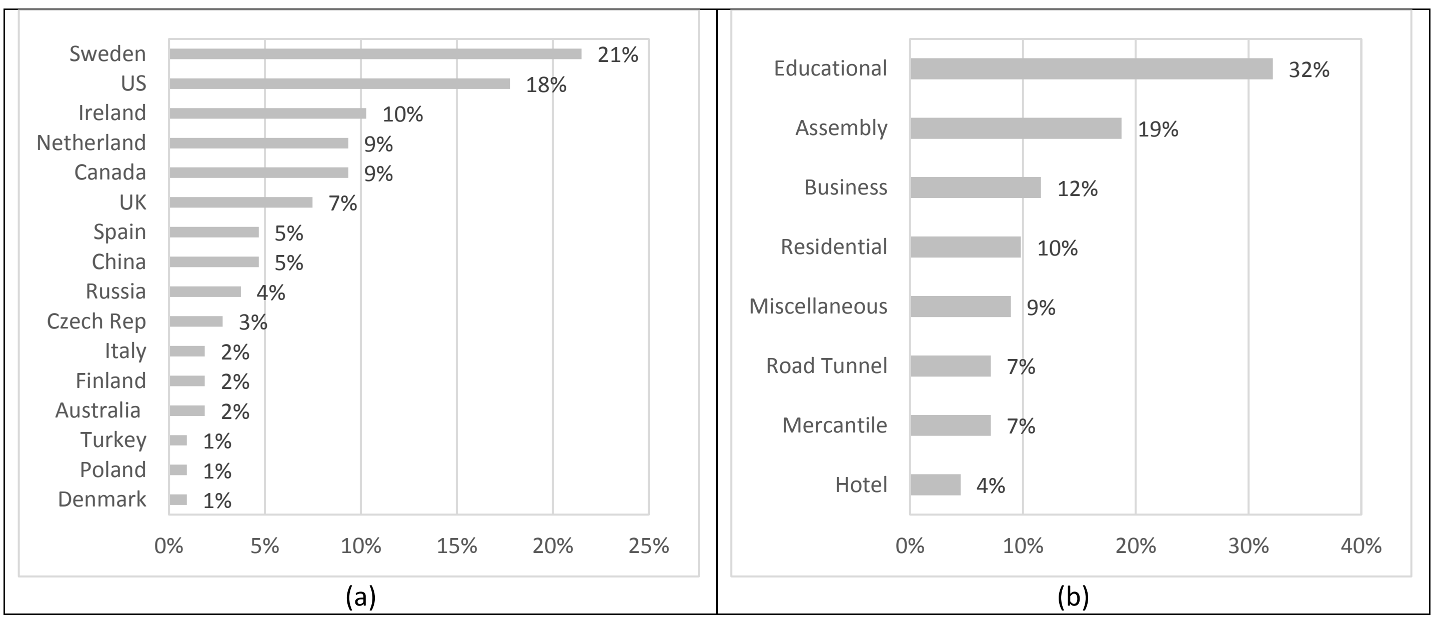

Figure 3 – Percentages of the case studies for each (a) country and (b) occupancy

[5] Such a comparison can be only carried out using the mean values of each case study as the standard deviations are not reported by Fahy and Proulx [28] and Shi et al. [27].

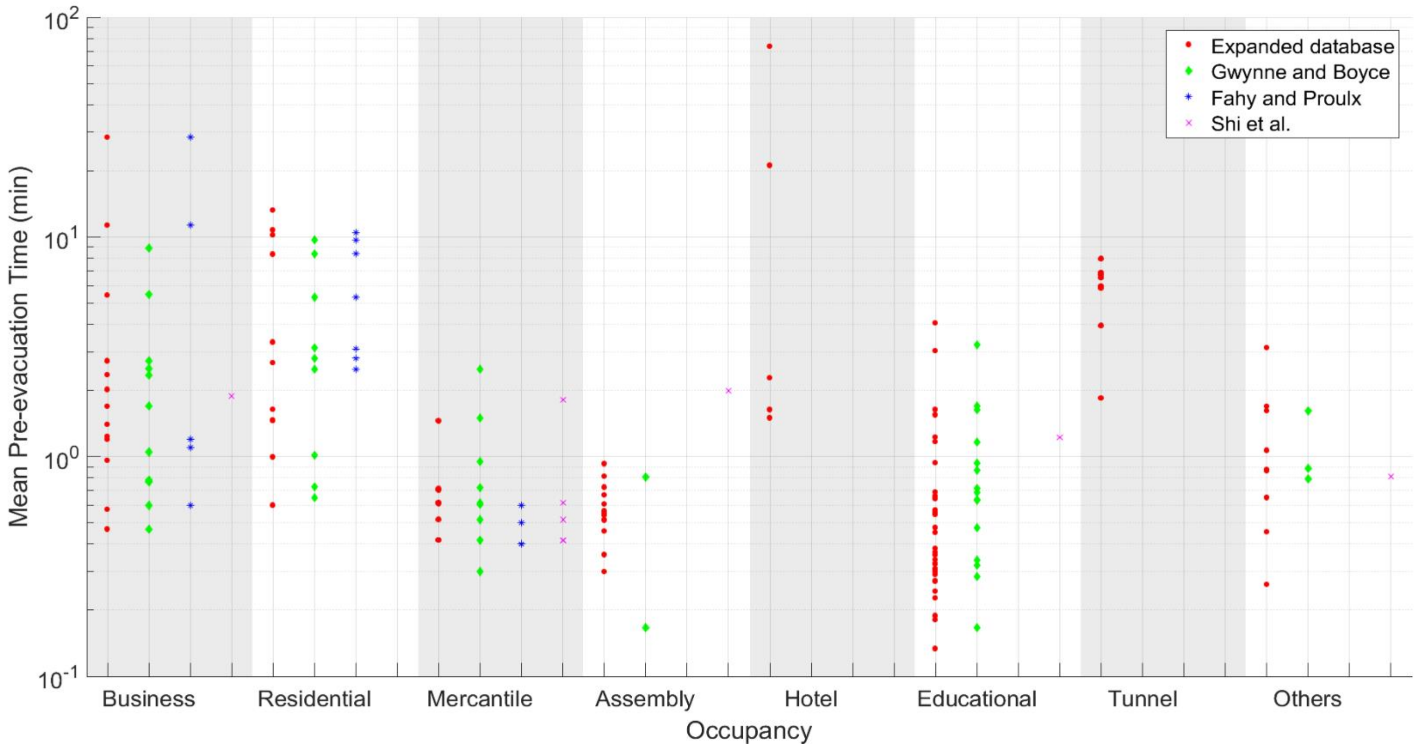

Figure 4 – Comparison between the proposed database with the database proposed by Gwynne and Boyce [4], Fahy and Proulx [28] and Shi et al. [27]

Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it was possible to identify three clusters using the k-mean cluster analysis (Section 3.2) as illustrated in Figure 7. From Figure 7, it is possible to observe that Clusters 2 and 3 represent four extreme case studies having the greatest pre-evacuation times. Those case studies are all fire incidents which took place in a high rise hotel and office buildings, listed in Table 1.

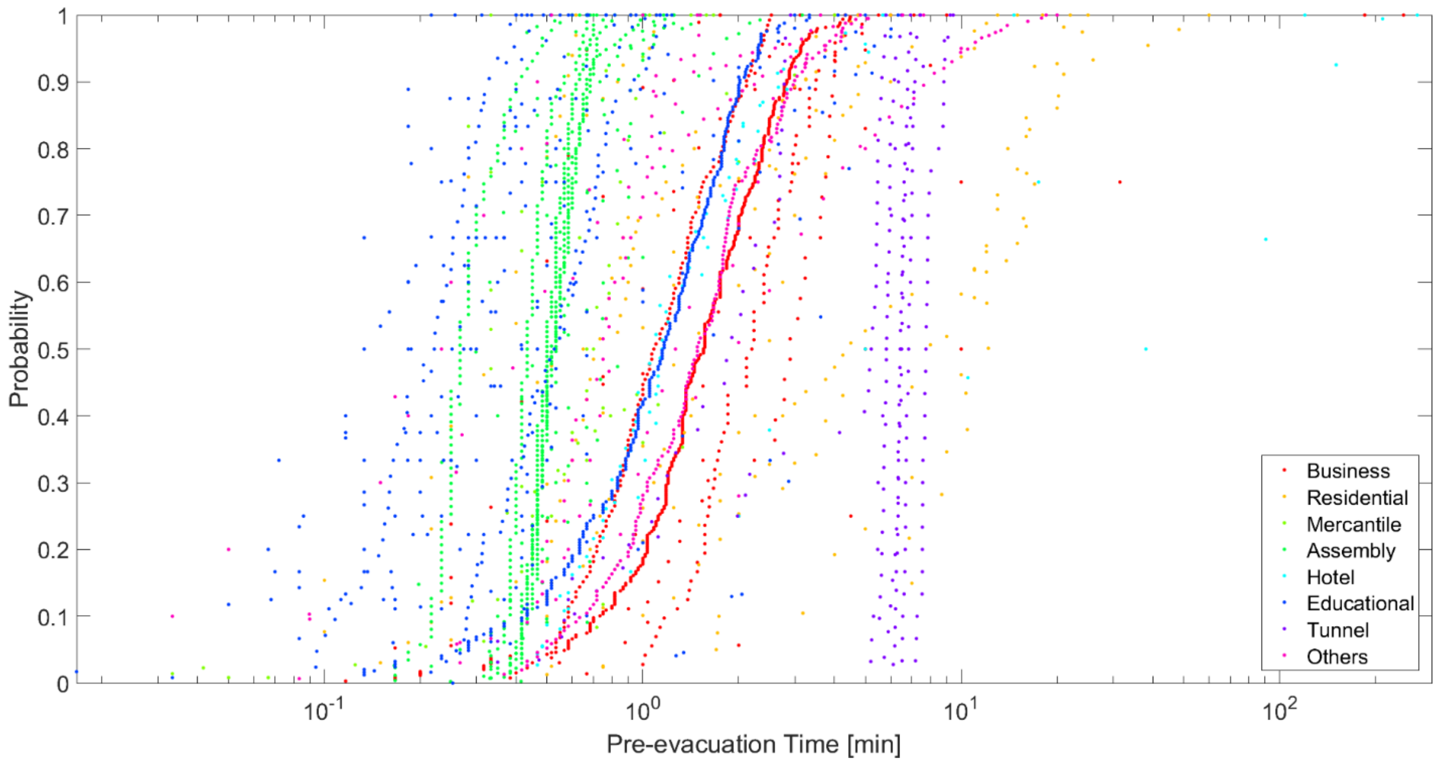

Figure 5 – Available data points by occupancy

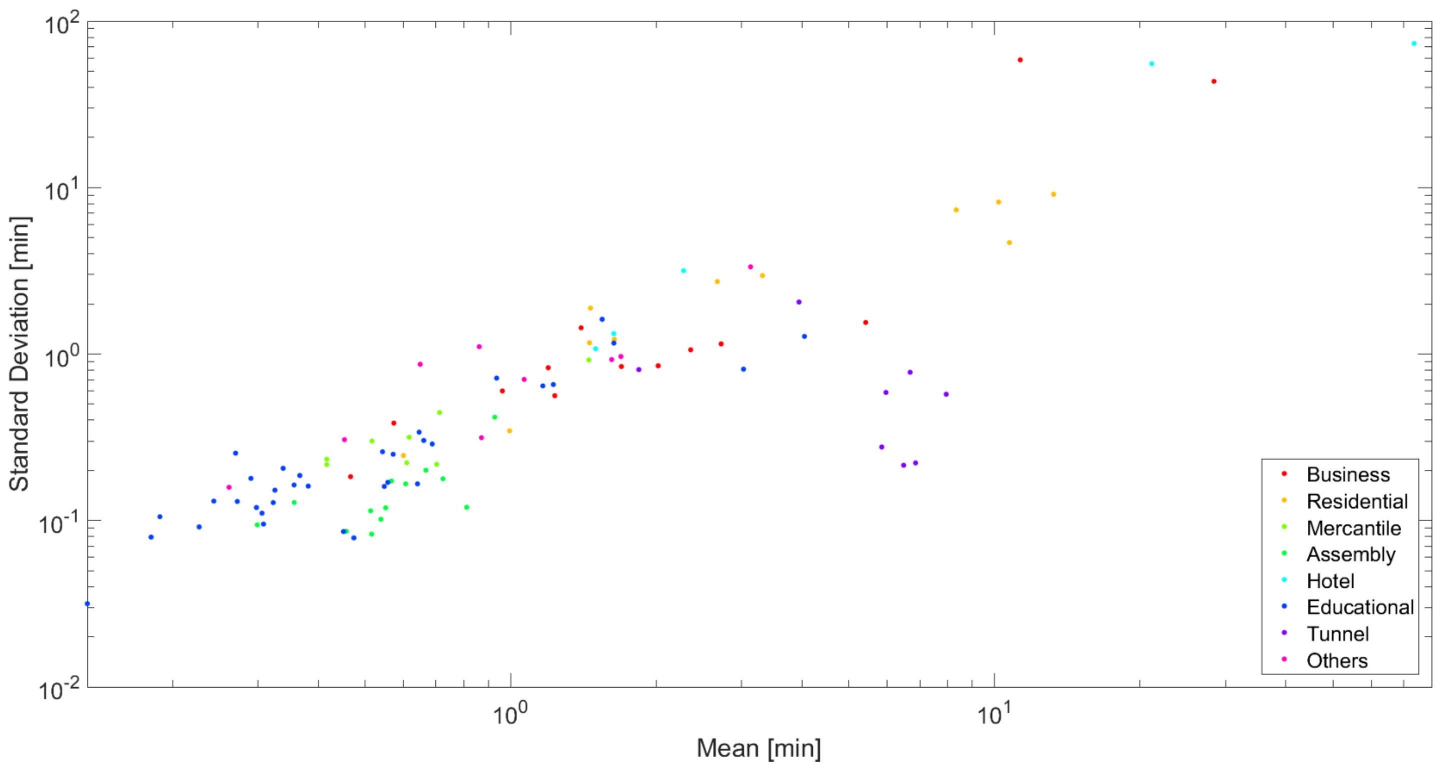

Figure 6 – Mean and standard deviation of the case study divided by occupancy

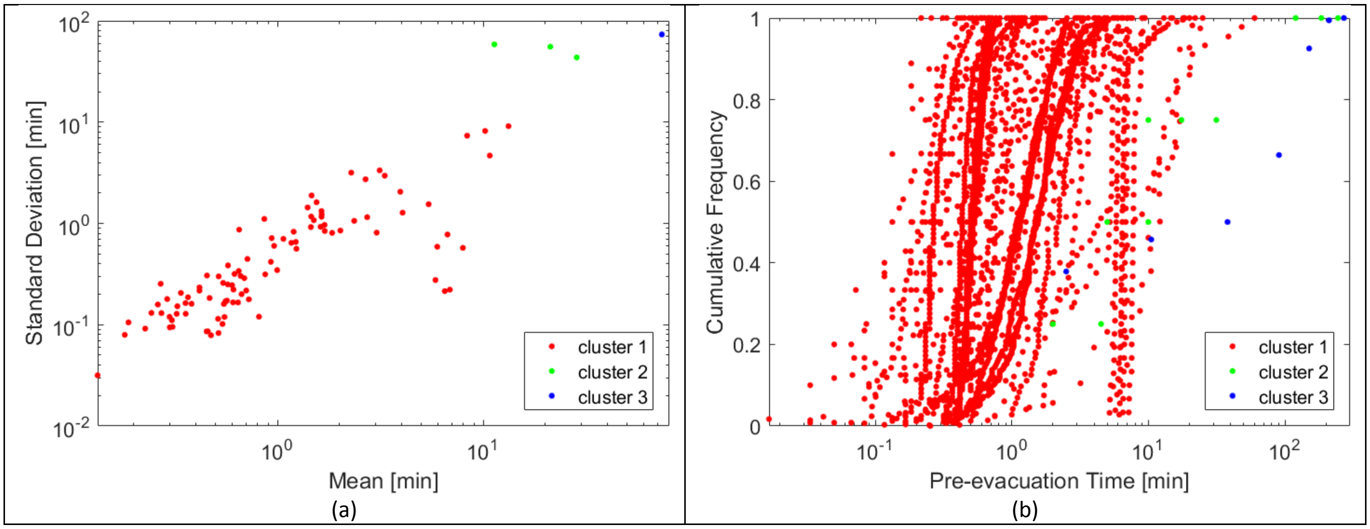

Figure 7– Pre-evacuation data divided by clusters: (a) means and standard deviations; (b) data points.

Table 1 – Pre-evacuation data for the four case studies having the greatest mean pre-evacuation time and standard deviations.

Ref.	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[28]	High-rise office	US	FI	No AL	110	85	11.300	58.489	2
[28]	High-rise office	US	FI	No AL	110	46	28.400	43.490	2
[38]	High-rise hotel	US	FI	No AL	13	47	21.125	55.510	2
[39]	High-rise hotel	US	FI	No AL	26	536	73.613	73.370	3

In the following subsections, the pre-evacuation data is analyzed according to each occupancy type. For each occupancy type, the case studies and characteristics (noted in Section 3.1 above) are provided in tabular format. These case studies are clustered, when possible, using the approach introduced in Section 3.3. For each occupancy an attempt can be made to explain/interpret the clustering. In some cases, this may be obvious given the characteristics of the case studies presented in the tables; in other the clustering may be rather difficult to explain with the reasons being much more complex.

Several factors can explain the clustering results. Some factors are reported in the following tables allowing the reader to infer any relationships; however, other influential factors may exist but are not included as they were not reported in the original referenced material. Any relationships drawn should then be considered provisional. The factors reported in the following tables are those selected in the SFPE Chapter [4] while the remaining are from the existing literature on pre-evacuation behavior. As such, we provide a comprehensive list of factors that readers can use to assess the difference among clusters:

1) The presence of an alarm, the type of alarm system and its performance [4];
2) Country and evacuee culture [4];
3) Nature of the event, i.e., fire accidents vs drills [4];
4) Type of building structure, e.g., number of floors, geometry, etc. [4];
5) Evacuation procedure [4];
6) Length of voice message and nature of the provided message [40,41];
7) Time of the day [42,43];
8) Weather conditions [44–46];

9) Difference in the methodology to collect behavioral data, such as closed-circuit television video analysis, questionnaires and interview [19,26];
10) Percentage of disabilities, elderly and motion impaired occupants [47].

Given the inconsistencies and omissions in the original 112 datasets, it was not possible for the authors to definitively establish the underlying factors that generated the clusters identified. As such, we suggest evacuation model users select the occupancy and the cluster that have the most similarities with their own case study depending on the list of those ten factors. The clusters might then act as a means by which to narrow the search of the data available. The readers should refer to the original references in case of uncertainty regarding the best cluster to select.

In Sections 4.1 – 4.8, we estimate four pre-evacuation distributions (i.e., gamma, lognormal, loglogistic and Weibull distributions) for each cluster. In choosing the most representative distribution, evacuation model users should consider both the R^2 parameters as well as the regression charts within the same cluster. In fact, the R^2 parameters provide an overall assessment of the fitting while the regression charts help identify distributions which fit better across the range of interest. Where users cannot use a particular distribution (e.g., where their evacuation model cannot represent a distribution), they can select an alternative distribution between the remaining ones. In both instances, the user will have an idea of how representative the curve is of the underlying data given the associated R2 value.

4.1 Business Occupancy

There were 13 case studies belonging to the business occupancy group. Those evacuations took place in buildings containing 4 to 110 floors located mostly in the US and Canada (50%). Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it was possible to identify two clusters as illustrated in Table 2. Figure 8.a illustrates the location of 13 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 8.b.

Table 2 – Pre-evacuation data for business occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[44,45]	1	Office	US	P-UD	PV	11	72	2.355	1.060	1
[44,45]	2	Mixed Office	US	UD	T3	4	348	1.693	0.841	1
[44,45]	3	Mixed Office	US	P-UD	PV	12	132	1.233	0.562	1
[48]	4	Office	Canada	UD	AL	13	458	1.398	1.436	1
[49]	5	Office	Canada	UD	AL	6	92	0.573	0.385	1
[49]	6	Office	Canada	UD	AL	7	161	1.196	0.827	1
[50]	7	Office*	Finland	AD	AL	7	33	2.722	1.151	1
[50]	8	Office*	Finland	AD	AL	4	9	2.017	0.850	1
[51]	9	Office	UK	UD	AL	6	19	0.467	0.183	1
[52]	10	Office	Denmark	UD	PV	12	70	0.961	0.600	1
[42]	11	Office	Australia	FI	No AL	14	106	5.415	1.547	1
[28]	12	Office**	US	FI	No AL	110	85	11.300	58.489	2
[28]	13	Office**	US	FI	No AL	110	46	28.400	43.490	2

* Data converted from graphical to digital form
** World Trade Center

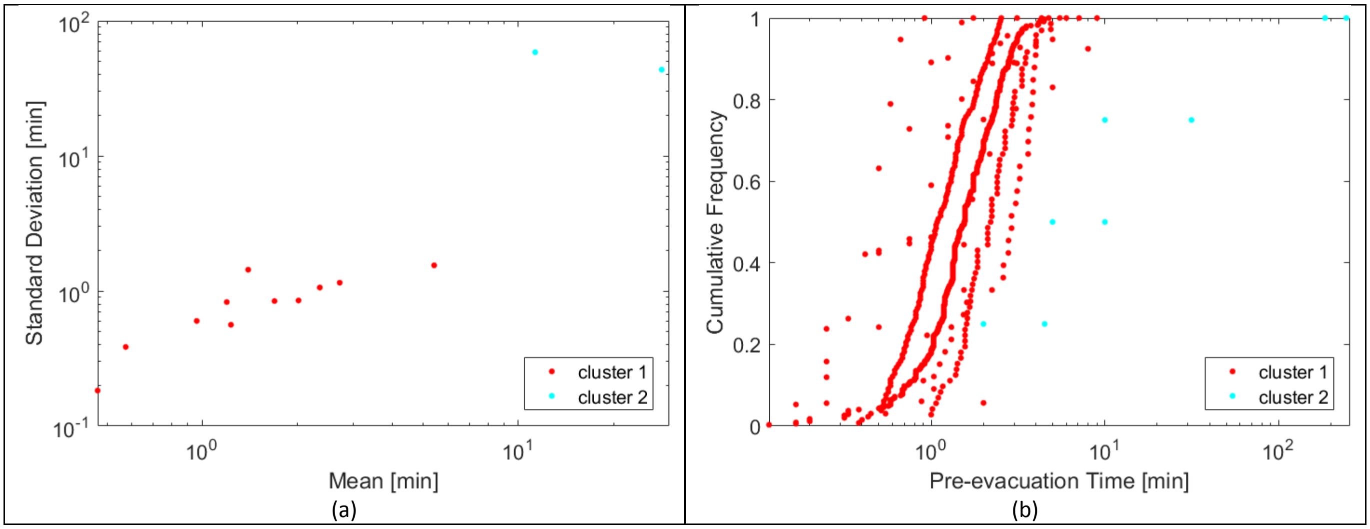

Figure 8 – Pre-evacuation data for business occupancy by clusters: (a) means and standard divisions of the case studies; (b) data point of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimate parameters and the R^2 are displayed in Table 3. Those distributions and the related data points are illustrated in Figure 9.

Table 3 – Estimated parameters of the pre-evacuation distributions for the business clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma	1.291	103.901	1.291	1.732	2597	0.564
	Lognormal	381.651	0.967	40.919	0.967		0.548
	Loglogistic	4.592	0.587	0.498	0.587		0.548
	Weibull	139.285	1.195	2.321	1.195		0.566
2	Gamma	0.557	1419.096	0.557	23.651	10	0.942
	Lognormal	36.131	1.613	11.104	1.613		0.949
	Loglogistic	5.905	0.958	1.811	0.958		0.950
	Weibull	672.010	0.664	11.200	0.664		0.944

It is worth highlighting that Cluster 2 refers to the evacuation that took place in the World Trade Center in 1993. For this evacuation, it was possible to find only ten percentiles (see Figure 8.b) from the literature. This small number of data points explains the high value of R^2 in Table 3. Regardless of the limitation of this case study, the proposed distribution is capable of representing the pre-evacuation event as illustrated in Figure 9.

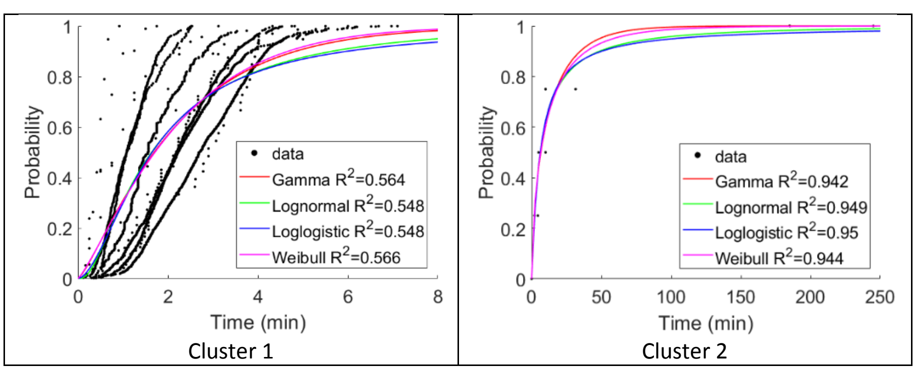

Figure 9 – Regression charts for the business clusters

4.2 Residential Occupancy

There were 11 residential occupancy case studies. Those evacuations took place in buildings containing 4 to 30 floors located mostly in Canada (63%). For this occupancy, we exclude the sleep data presented in [4] as they focus on the effectiveness of different notification systems to wake sleeping participants and on arousal time rather than pre-evacuation time. Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it is possible to identify two clusters as illustrated in Table 4. Figure 10.a illustrates the location of 11 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 10.b.

Table 4 – Pre-evacuation data for residential occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[45]	1	Univ. Residence	US	UD	T3	4	40	0.995	0.346	1
[45]	2	Univ. Residence	US	UD	T3	4	33	0.600	0.245	1
[53]	3	Apartment	Canada	UD	AL	7	42	2.673	2.727	1
[53]	4	Apartment	Canada	UD	AL	7	80	3.313	2.961	1
[54]	5	Apartment	Canada	UD	AL	14	31	1.637	1.225	1
[54]	6	Apartment	Canada	UD	AL	14	94	1.455	1.169	1
[47]	7	Residential Care	UK	UD	AL	3	13	1.462	1.887	1
[43]	8	Apartment*	Canada	FI	AL[a b]	30	103	13.239	9.127	2
[42]	9	Apartment	Australia	FI	AL[a]	18	26	10.731	4.670	2
[53]	10	Apartment	Canada	UD	AL[b]	6	55	8.332	7.344	2
[53]	11	Apartment	Canada	UD	AL[b]	7	79	10.190	8.182	2

* Data converted from graphical to digital form

[a] Early morning; [b] Poor performance of the alarm system

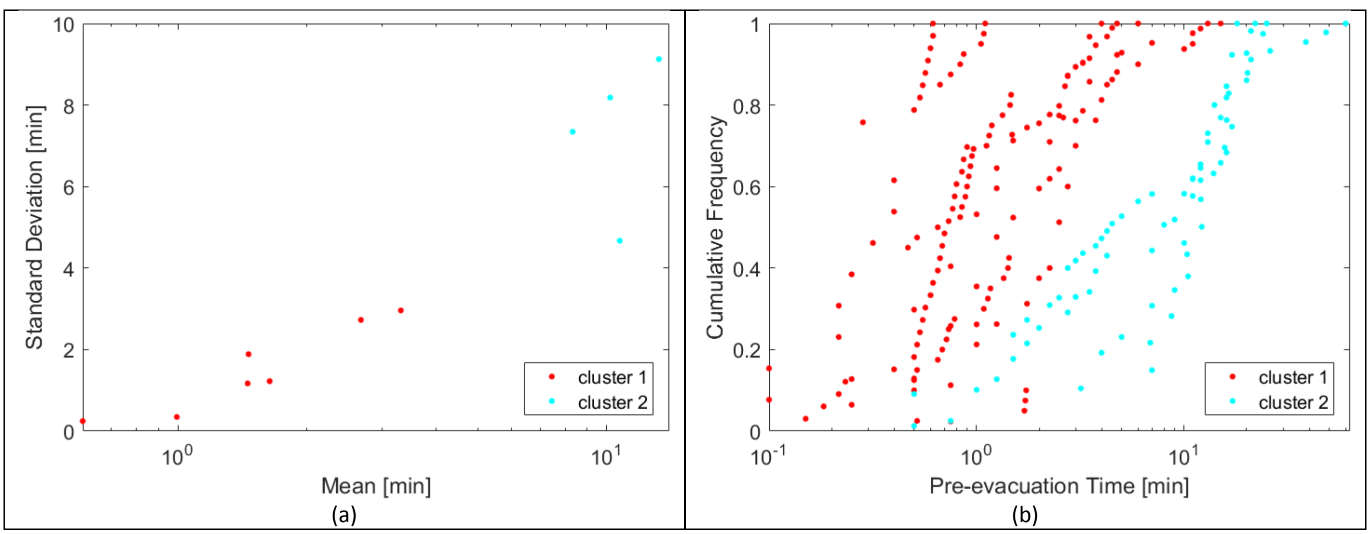

Figure 10 – Pre-evacuation data for residential occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data points of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimate parameters and the R^2 are displayed in Table 5. Those distributions and the related data points are illustrated in Figure 11.

Table 5 – Estimate parameters of the pre-evacuation distributions for the residential clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma	0.650	178.024	0.650	2.967	149	0.601
	Lognormal	54.879	1.432	-0.119	1.432		0.589
	Loglogistic	4.087	0.873	-0.007	0.873		0.586
	Weibull	102.475	0.767	1.708	0.767		0.599
2	Gamma	0.911	812.708	0.911	13.545	78	0.820
	Lognormal	98.986	1.268	32.821	1.268		0.785
	Loglogistic	6.143	0.763	2.049	0.763		0.784
	Weibull	724.617	0.978	12.077	0.978		0.819

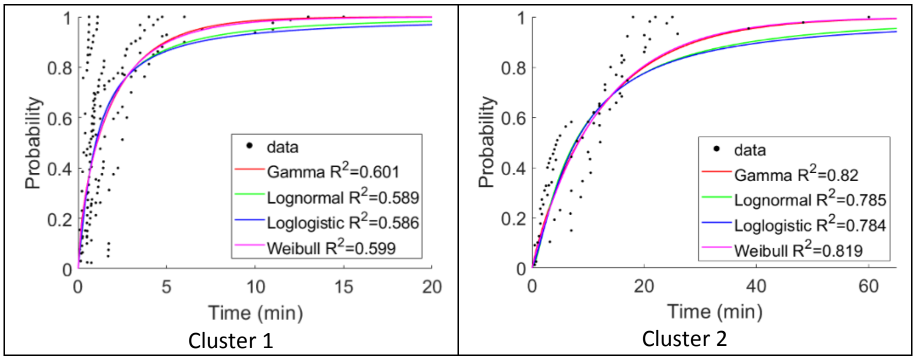

Figure 11 – Regression charts for the residential clusters

The results In Table 5 and Figure 11 show a reasonably good agreement between existing data and proposed distributions. In this case, the lower values of R^2 for Cluster 1 can be explained by the high dispersion of the data points of this cluster as shown in Figure 11. However, given the regression charts in Figure 11 the proposed distributions seem to provide a good representation of the trends shown by the data points.

4.3 Mercantile Occupancy

There were 8 case studies belonging to this occupancy group. Those evacuations took place in buildings containing 1 to 3 floors located mostly in the UK and Sweden (88%). Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it is possible to identify two clusters as illustrated in Table 6. Figure 12.a illustrates the location of the 8 case studies on the mean vs. standard deviation plane while the data points of those case studies are displayed in Figure 12.b.

Table 6 – Pre-evacuation data for mercantile occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[55]	1	Marks & Spencer (Sprucefield Centre)	UK	UD	AL	1	95	0.417	0.233	1
[56]	2	Marks & Spencer (Culverhouse Cross)	UK	UD	AL	1	71	0.417	0.217	1
[57]	3	IKEA (Örebro)	Sweden	UD	VA	1	16	0.703	0.217	1
[57]	4	IKEA (Västerås)	Sweden	UD	VA	1	12	0.610	0.222	1
[57]	5	IKEA (Älmhult)	Sweden	UD	VA	3	17	0.713	0.445	1
[56]	6	Marks & Spencer (Royal Ave)	UK	UD	AL	3	122	0.617	0.317	1
[56]	7	Marks & Spencer (Queen St)	UK	UD	AL	3	122	0.517	0.300	1
[58]	8	Xin Lian Xin Store	China	AD	AL	1	294	1.450	0.921	2

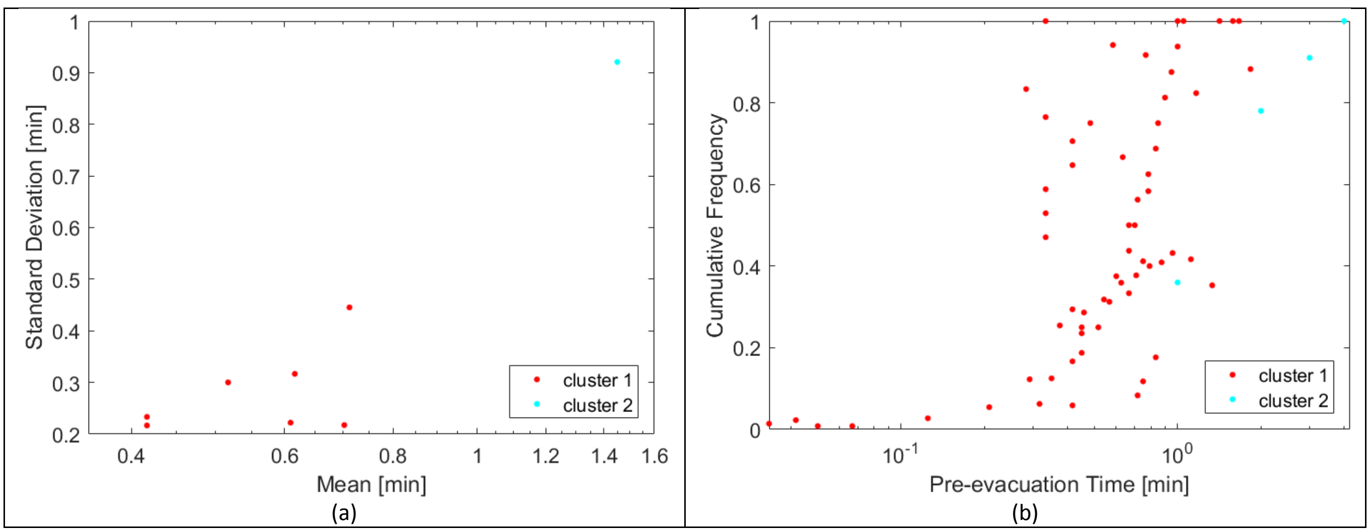

Figure 12 – Pre-evacuation data for mercantile occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data points of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimated parameters and the R^2 values are displayed in Table 7. Those distributions and the related data points are illustrated in Figure 13.

Table 7 – Estimate parameters of the pre-evacuation distributions for the mercantile clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a (s)	b (s)	a (min)	b (min)		
1	Gamma	3.005	14.564	3.005	0.243	4	0.901
	Lognormal	62.874	0.574	-7.764	0.574		0.895
	Loglogistic	3.660	0.334	-0.434	0.334		0.893
	Weibull	48.453	1.957	0.808	1.957		0.905
2	Gamma	2.535	34.561	2.535	0.576	63	0.996
	Lognormal	150.491	0.610	7.704	0.610		0.996
	Loglogistic	4.309	0.362	0.215	0.362		0.993
	Weibull	96.470	1.642	1.608	1.642		0.994

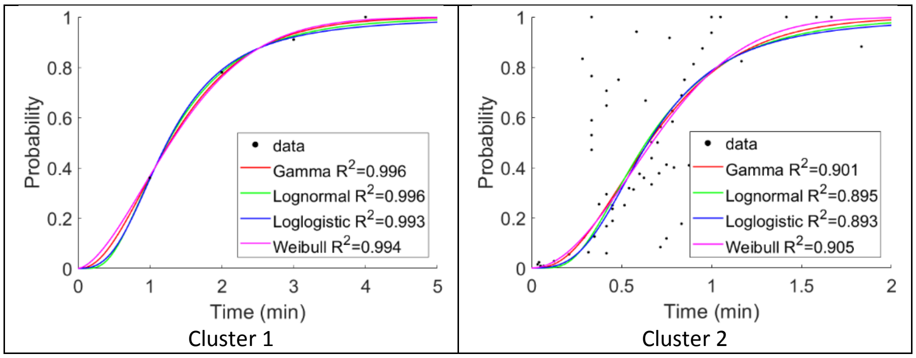

Figure 13 – Regression charts for the mercantile clusters

It is worth highlighting that Cluster 2 refers to the evacuation that took place in the Xin Lian Xin Store in China. For this evacuation, it was possible to find only four data points from the literature. This small number of data points explains the high value of R^2 in Table 7. Regardless of the limitation of this case study, the proposed distribution is capable of representing the pre-evacuation event as illustrated in Figure 13.

4.4 Assembly Occupancy

There were 21 case studies belonging to the assembly occupancy group. Those evacuations took place in buildings containing 1 to 3 floors located in the UK, Sweden, and China. Cluster analysis was conducted for the cinema and theatre case studies, only. Using the mean and the standard deviation of the pre-evacuation time of those cinema evacuations, it is possible to identify four clusters as illustrated in Table 8. Figure 14.a illustrates the location of 21 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 14.b. Considering the differences between buildings types within this specific occupancy category, we manually defined independent clusters for the restaurant/bar case studies. This occupancy contains a broader range of building types that may require different pre-evacuation data, e.g., restaurants/bars vs. cinema theaters. Since it is likely that engineers will search for data related to building type (even within the same occupancy), this approach simplifies the users' choice by manually separating those case studies from restaurants/bars and those from cinema theaters into separate clusters.

Table 8 – Pre-evacuation data for assembly occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[59]	1	Theatre*	UK	UD	PV	3	338	0.357	0.128	1
[60]	2	Cinema-Theatre	Sweden	UD	AL	1	87	0.299	0.094	1
[60]	3	Cinema-Theatre	Sweden	UD	AL	1	98	0.457	0.086	2
[60]	4	Cinema-Theatre	Sweden	UD	PV	1	108	0.539	0.102	2
[60]	5	Cinema-Theatre	Sweden	UD	PV	1	128	0.516	0.083	2
[60]	6	Cinema-Theatre	Sweden	UD	PV	1	129	0.551	0.119	2
[61]	7,8,9	Cinema*	Sweden	UD	PV	1	126	0.607	0.166	2
[61]	10,11,12	Cinema*	Sweden	UD	PV	1	297	0.567	0.173	2
[61]	13,14,15	Cinema*	Sweden	UD	PV	1	39	0.725	0.178	3
[61]	16,17,18	Cinema*	Sweden	UD	PV	1	178	0.668	0.200	3
[7]	19	Theatre*	UK	UD	LV+PV	3	115	0.927	0.418	4
[7]	20	Restaurant*	UK	UD	AL+PV	2	11	0.811	0.120	5
[62]	21	Bar	China	UD	-	1	40	0.513	0.114	5

* Data converted from graphical to digital form

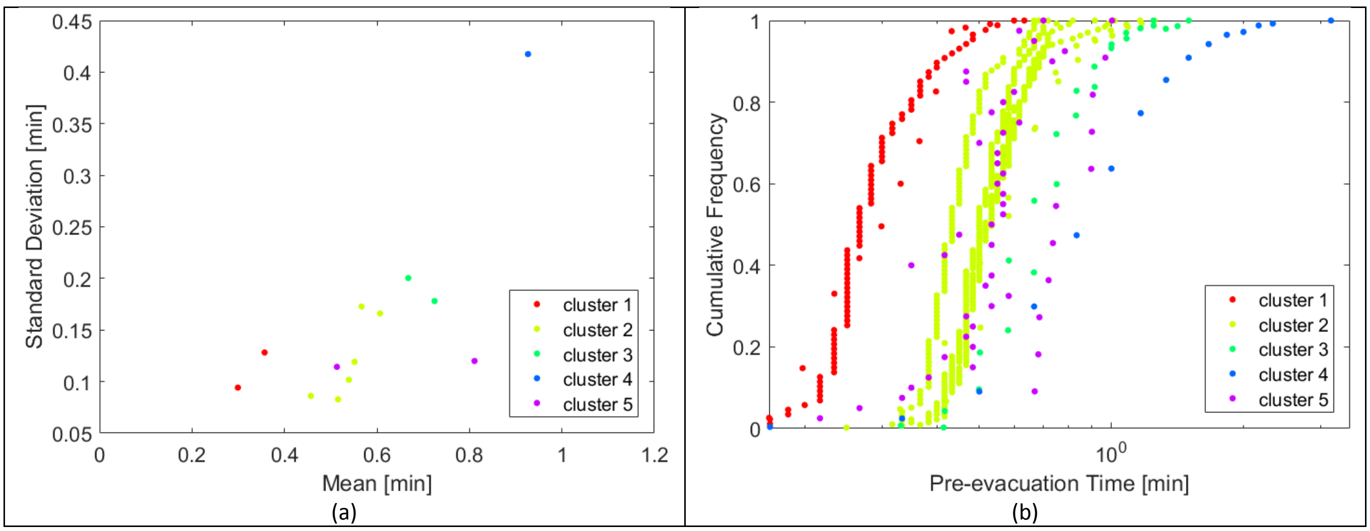

Figure 14 – Pre-evacuation data for assembly occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data point of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimated parameters and the R^2 are displayed in Table 9.

Table 9 – Estimate parameters of the pre-evacuation distributions for the assembly clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		R^2
		a	***b***	***a***	***b***	
1	Gamma	10.584	1.664	10.584	0.028	0.961
1	Lognormal	417.150	0.308	-186.262	0.308	0.962
1	Loglogistic	2.829	0.184	-1.266	0.184	0.960
1	Weibull	19.187	3.554	0.320	3.554	0.954
2	Gamma	16.627	1.898	16.627	0.032	0.886
2	Lognormal	798.053	0.248	-154.961	0.248	0.888
2	Loglogistic	3.428	0.149	-0.666	0.149	0.889
2	Weibull	33.790	4.631	0.563	4.631	0.880
3	Gamma	12.779	3.138	12.779	0.052	0.992
3	Lognormal	428.761	0.281	-51.083	0.281	0.993
3	Loglogistic	3.658	0.166	-0.436	0.166	0.992
3	Weibull	43.514	4.008	0.725	4.008	0.986
4	Gamma	5.544	9.949	5.544	0.166	0.999
4	Lognormal	904.832	0.427	-36.708	0.427	1.000
4	Loglogistic	3.935	0.251	-0.159	0.251	0.999
4	Weibull	60.959	2.579	1.016	2.579	0.996
5	Gamma	2.861	12.757	2.861	0.213	0.372
5	Lognormal	62.497	0.611	-11.308	0.611	0.378
5	Loglogistic	3.466	0.376	-0.628	0.376	0.377
5	Weibull	39.969	1.780	0.666	1.780	0.367

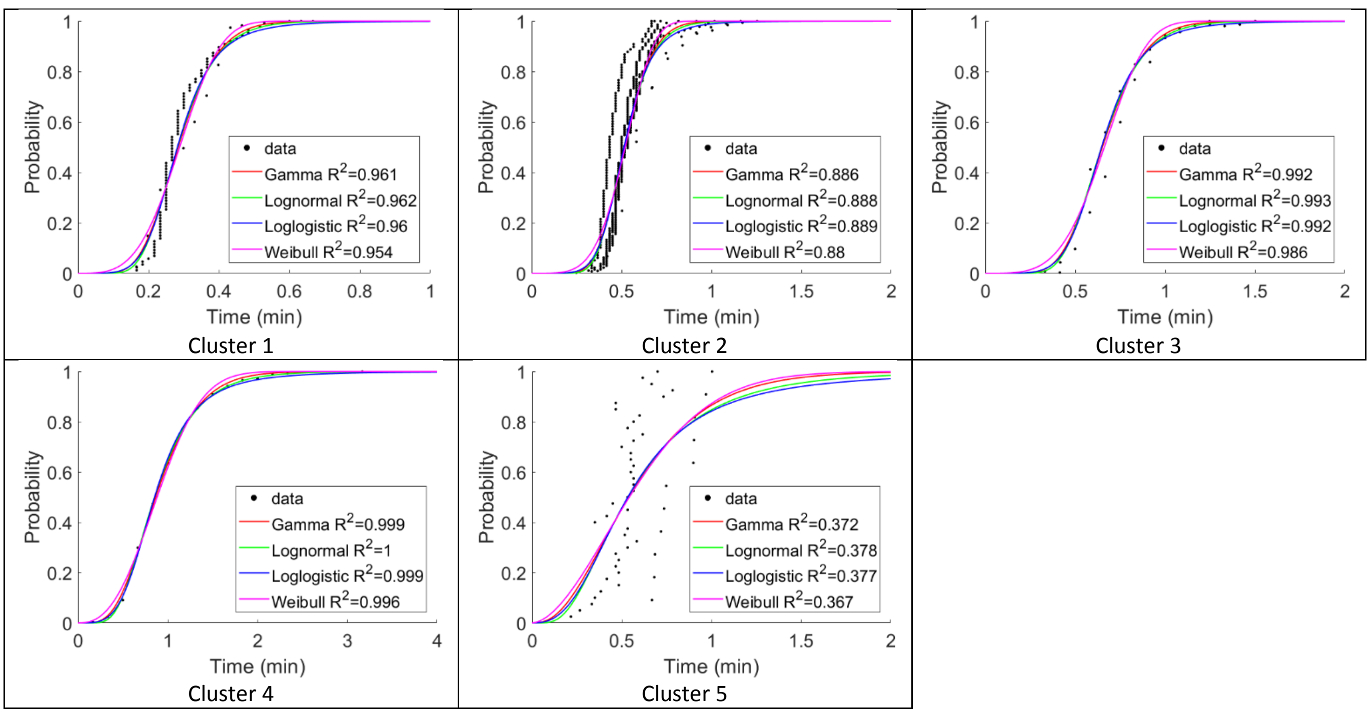

Figure 15 – Regression charts for the assembly clusters

The results in Table 9 and Figure 15 show good agreement between the existing data and the proposed distributions except for Cluster 5. This is due to the high dispersion of the data from the two evacuations which took place in a bar and a restaurant. For Cluster 4, it is possible to observe a close match between data and distribution (i.e., the R^2 parameter is very close to one). This result can be explained by the fact that the 15 data points (see Figure 14.b) were from a single evacuation from a theatre in UK.

4.5 Hotel Occupancy

There were five case studies belonging to the hotel occupancy group. Those evacuations took place in buildings located mostly in the Netherlands and the US. Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it is possible to identify three clusters as illustrated in Table 10. Figure 16.a illustrates the location of 5 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 16.b.

Table 10 – Pre-evacuation data for hotel occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floors	Sample	Mean [min]	SD [min]	Cluster
[63]	1	Hotel	Netherland	UD	Phone Message	**	18	2.277	3.169	1
[63]	2	Hotel	Netherland	UD	Phone Message	**	37	1.498	1.076	1
[63]	3	Hotel	Netherland	UD	Phone Message	**	23	1.633	1.325	1
[38]	4	High-rise hotel	US	FI	No Alarm	13	47	21.125	55.510	2
[39]	5	High-rise hotel	US	FI	No Alarm	26	536	73.613	73.370	3

** those experiments took place on a single floor and corridor

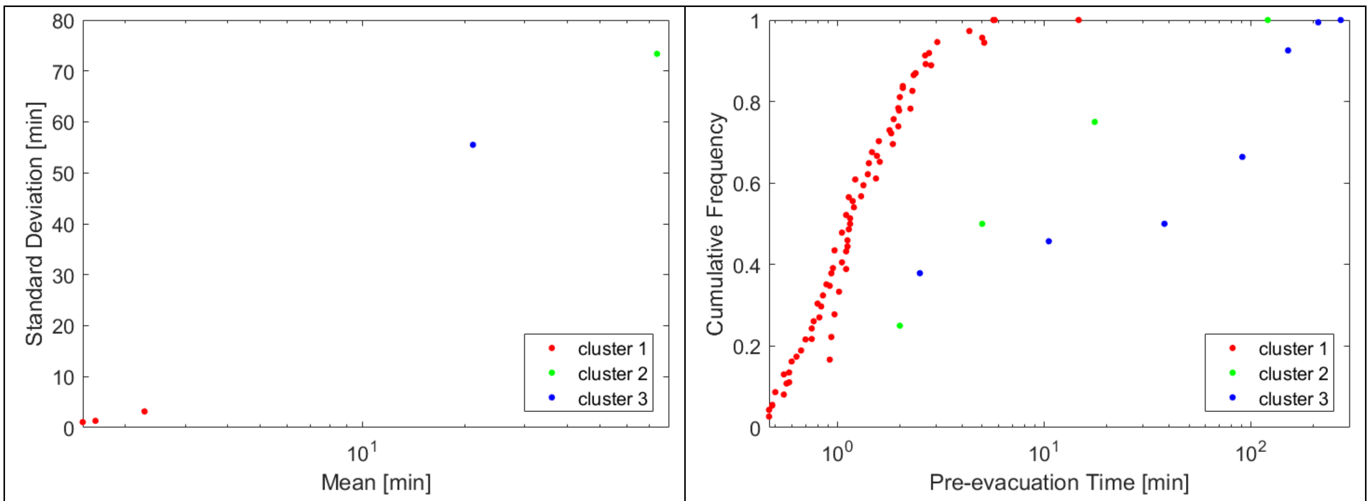

Figure 16 – Pre-evacuation data for hotel occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data point of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimate parameters and the R^2 are displayed in Table 11. Those distributions and the related data points are illustrated in Figure 17.

Table 11 – Estimate parameters of the pre-evacuation distributions for the hotel clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma	2.787	29.503	2.787	0.492	78	0.974
	Lognormal	423.443	0.631	16.852	0.631		0.978
	Loglogistic	4.262	0.379	0.168	0.379		0.978
	Weibull	90.500	1.790	1.508	1.790		0.968
2	Gamma	0.567	1276.836	0.567	21.281	4	0.994
	Lognormal	93.846	1.571	27.487	1.571		0.997
	Loglogistic	5.780	0.946	1.686	0.946		0.996
	Weibull	606.399	0.684	10.107	0.684		0.996
3	Gamma	0.294	14165.384	0.294	236.085	7	0.875
	Lognormal	13.082	2.560	4.947	2.560		0.784
	Loglogistic	6.589	1.567	2.495	1.567		0.772
	Weibull	2088.400	0.440	34.806	0.440		0.835

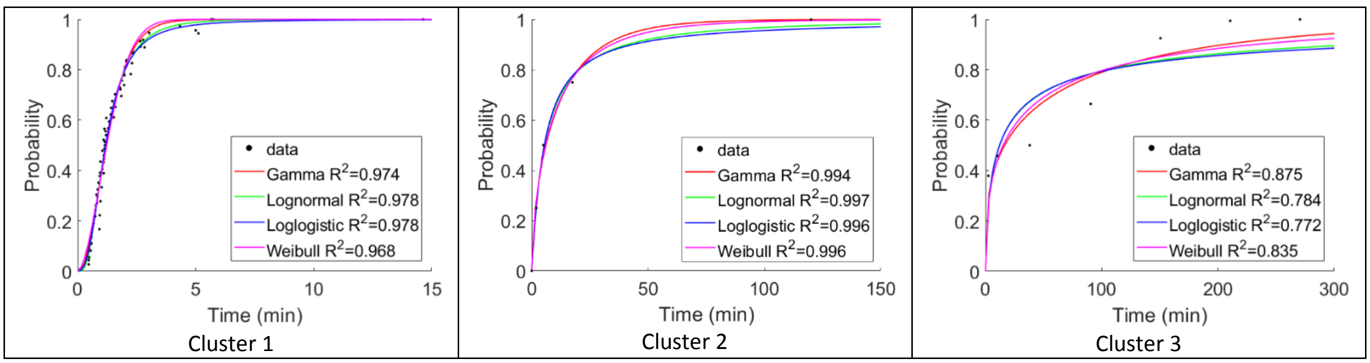

Figure 17 – Regression charts for the hotel clusters

The results In Table 11 and Figure 17 show good agreement between the existing data and the proposed distributions. It is worth highlighting that Cluster 2 refers to the evacuation that took place in a high-rise hotel in the US. For this evacuation, it was possible to find only five data points from the literature. This small number of data points explains the high value of R^2 in Table 11. Regardless of the limitations of this case study, the proposed distribution is capable representing the pre-evacuation event as illustrated in Figure 17.

4.6 Educational Occupancy

There were 36 educational case studies. These took place in schools, libraries, laboratories and university lecture halls. A cluster analysis is used to analyze the case studies of kindergartens, pre-schools, primary and secondary school buildings (i.e., case studies 1-22). Using the mean and the standard deviation of the pre-evacuation time of those evacuations, it is possible to identify two clusters as illustrated in Table 12. It is worth highlighting that the pre-evacuation times of schools, i.e., case studies 7-22, refer to the classroom instead of single evacuees as the students belonging to the same classroom evacuate as a single group.

The remaining case studies are manually divided (without the use of cluster analysis) accordingly to the type of buildings into clusters referring to Library (Clusters 3), Laboratory (Cluster 4) and Lecture hall (Cluster 5) as illustrated in Table 12. This occupancy contains a broader range of building types that may require different pre-evacuation data. Since it is likely that engineers will search for data related to building type (even within the same occupancy), this approach simplifies the users' choice by manually separating those case studies.

Figure 18.a illustrates the location of the 26 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 18.b.

Table 12 – Pre-evacuation data for educational occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Floor	Sample	Mean [min]	SD [min]	Cluster
[64]	1	Kindergarten*[a]	Russia	UD	AL	>1	25	3.027	0.811	1
[64]	2	Kindergarten*[b]	Russia	UD	AL	>1	77	4.046	1.276	1
[64]	3	Kindergarten*[c]	Russia	UD	AL	>1	52	0.451	0.086	2
[64]	4	Kindergarten*[d]	Russia	UD	AL	>1	34	0.642	0.166	2
[65]	5	Pre-school	Czech Rep.	SAD	Verbal	3	106	0.647	0.339	2
[65]	6	Pre-school	Czech Rep.	SAD	Verbal	3	101	0.543	0.259	2
[66]	7	Primary School-1*	Ireland	UD	AL	2	228	0.557	0.169	2
[66]	8	Primary School-1*	Ireland	UD	AL	2	210	0.357	0.163	2
[66]	9	Primary School-1*	Ireland	UD	AL	2	234	0.306	0.111	2
[66]	10	Primary School-2*	Ireland	UD	AL	2	263	0.133	0.032	2
[66]	11	Primary School-2*	Ireland	UD	AL	2	268	0.339	0.206	2
[66]	12	Primary School-2*	Ireland	UD	AL	2	259	0.181	0.079	2
[66]	13	Primary School-3*	Ireland	UD	AL	2	144	0.290	0.179	2
[66]	14	Primary School-3*	Ireland	UD	AL	2	140	0.243	0.131	2
[66]	15	Primary School-4*	Ireland	UD	AL	2	195	0.366	0.186	2
[66]	16	Primary School-4*	Ireland	UD	AL	2	187	0.326	0.152	2
[66]	17	Primary School-4*	Ireland	UD	AL	2	170	0.298	0.120	2
[67]	18	Primary and secondary school	Spain	SAD	AL	3	131	0.270	0.254	2
[67]	19	Primary and secondary school	Spain	UD	AL	3	167	0.308	0.095	2
[67]	20	Primary and secondary school	Spain	UD	AL	3	247	0.661	0.302	2
[67]	21	Primary and secondary school	Spain	UD	AL	3	244	0.381	0.161	2
[67]	22	Primary and secondary school	Spain	UD	AL	3	243	0.323	0.128	2
[68]	23	Library	Poland	UD	PV	3	192	1.165	0.644	3
[69]	24	Library	Turkey	UD	AL + PV	2	51	0.935	0.716	3
[70]	25	Library*	Czech Rep.	UD	AL+ PV +LM	2	70	1.545	1.617	3
[71]	26	Library	UK	UD	AL	3	119	1.633	1.164	3
[72]	27	Library	UK	UD	AL	3	247	1.225	0.656	3
[40]	28	Laboratory	UK	UD	PV	**	17	0.688	0.288	4
[40]	29	Laboratory	UK	UD	PV	**	16	0.474	0.079	4
[40]	30	Laboratory	UK	UD	PV	**	15	0.272	0.130	4
[73]	31	Lecture hall	China	AD	AL	**	60	0.188	0.105	5
[74]	32	Lecture hall	Italy	UD	AL	**	62	0.572	0.250	5
[74]	33	Lecture hall	Italy	UD	AL	**	42	0.227	0.091	5
[61]	34,35,36	Lecture hall	Sweden	UD	PV	**	187	0.548	0.160	5

* Data converted from graphical to digital form

[a] Autumn or Spring; [b] Winter; [c] Summer; [d] Winter with blankets

** those experiments took place on single rooms

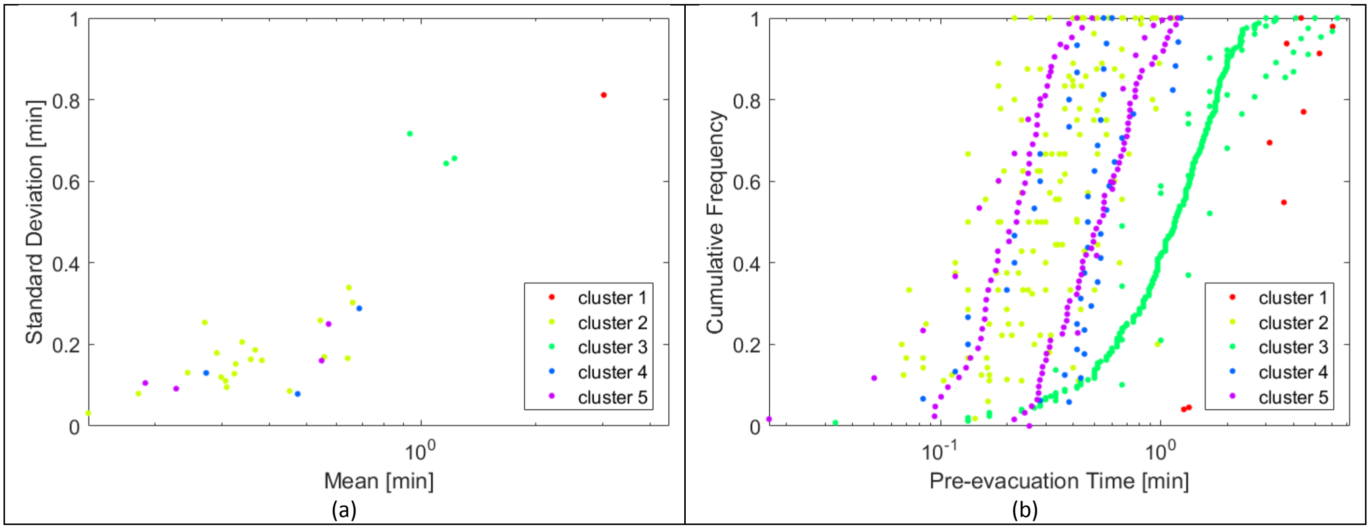

Figure 18 – Pre-evacuation data for educational occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data point of the case studies.

Four pre-evacuation distributions are estimated for each cluster. The estimate parameters and the R^2 are displayed in Table 13. Those distributions and the related data points are illustrated in Figure 19.

Table 13 – Estimate parameters of the pre-evacuation distributions for the educational clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma	7.859	24.692	7.859	0.412	14	0.931
	Lognormal	125.572	0.358	27.017	0.358		0.931
	Loglogistic	5.218	0.211	1.124	0.211		0.931
	Weibull	212.216	3.142	3.537	3.142		0.929
2	Gamma	0.757	30.111	0.757	0.502	141	0.377
	Lognormal	34.794	1.294	-20.181	1.294		0.380
	Loglogistic	2.591	0.794	-1.504	0.794		0.380
	Weibull	21.378	0.835	0.356	0.836		0.378
3	Gamma	2.743	28.042	2.743	0.467	291	0.937
	Lognormal	364.484	0.624	8.701	0.624		0.934
	Loglogistic	4.197	0.373	0.103	0.373		0.935
	Weibull	84.428	1.798	1.407	1.798		0.937
4	Gamma	1.117	30.385	1.118	0.506	48	0.386
	Lognormal	29.943	1.108	-8.731	1.108		0.372
	Loglogistic	3.177	0.675	-0.917	0.675		0.371
	Weibull	34.413	1.100	0.574	1.100		0.386
5	Gamma	1.144	24.612	1.144	0.410	127	0.528
	Lognormal	40.702	1.015	-15.379	1.015		0.504
	Loglogistic	2.977	0.620	-1.117	0.620		0.501
	Weibull	28.908	1.127	0.482	1.127		0.530

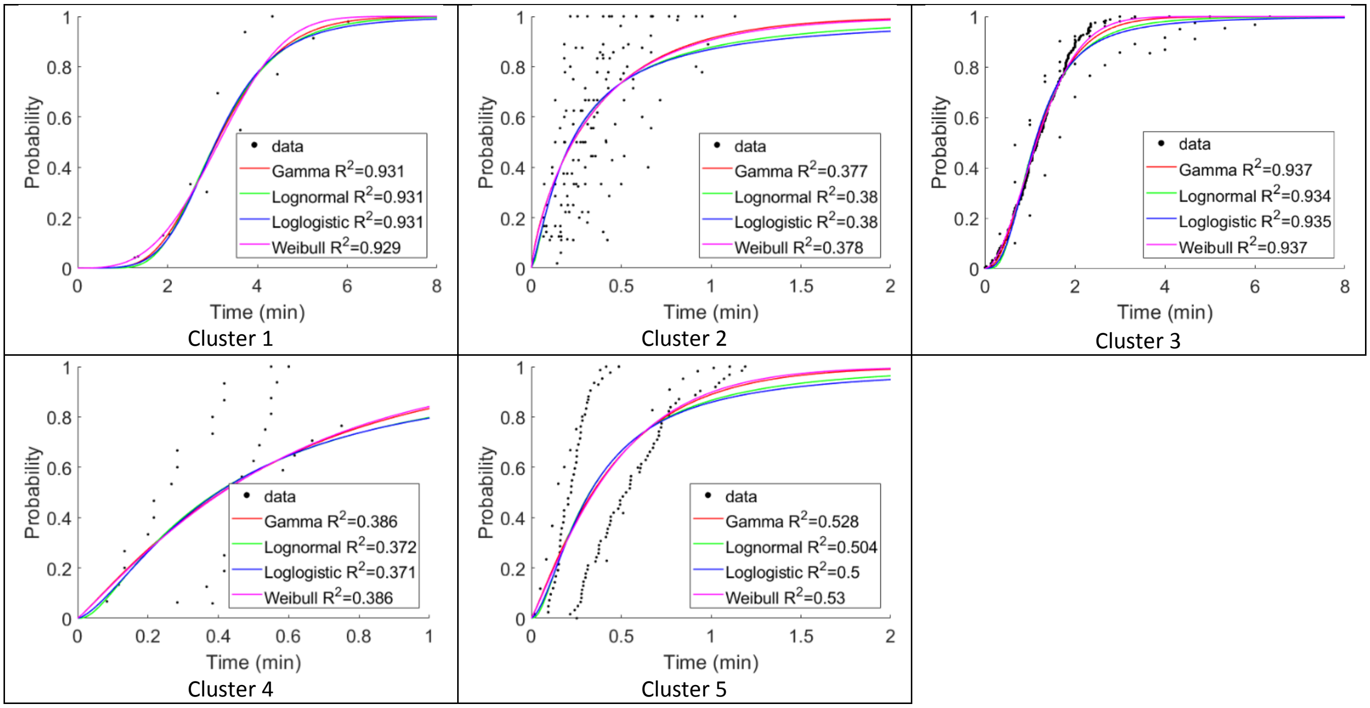

Figure 19 – Regression charts for the educational clusters

The results In Table 13 and Figure 19 show good agreement between the existing data and the proposed distributions for Cluster 1 and 3. The remaining clusters have a low value of R^2 as there is a large dispersion of the data belonging to those clusters. Moreover, Figure 19 indicates that the estimated distributions proposed for those remaining clusters are capable of representing the overall trends, but they do present some issues with the tails. As such the readers may need to use truncated versions of those distributions to account for this limitation.

4.7 Road Tunnel Occupancy

There are 8 case studies road tunnel case studies. Those evacuations took place in tunnels located in Sweden and Netherland. Again, it is possible to identify three clusters as illustrated in Table 14. Figure 20.a illustrates the location of 8 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 20.b.

Four pre-evacuation distributions are estimated for each cluster. In the case studies belonging to Cluster 3, all the evacuees evacuate after the alarm. Considering that the alarm was given at different times, we have normalized those data assuming that the time is equal to zero when the alarm goes off. The estimate the parameters and the R^2 are displayed in Table 15. Those distributions and the related data points are illustrated in Figure 21.

Table 14 – Pre-evacuation data for road tunnel occupancy

Ref.	Case Study	Building	Country	Nature	Alarm	Sample	Mean [min]	SD [min]	Cluster
[75]	1	Road Tunnel*	Sweden	UD	PV	29	1.840	0.806	1
[76]	2	Road Tunnel	Netherlands	UD	late PV (5min)	10	3.942	2.054	2
[76]	3	Road Tunnel	Netherlands	UD	late PV (4.7min)	31	5.845	0.276	3
[76]	4	Road Tunnel	Netherlands	UD	late PV (5.75min)	26	6.688	0.777	3
[76]	5	Road Tunnel	Netherlands	UD	late PV (5 min)	10	5.963	0.588	3
[76]	6	Road Tunnel	Netherlands	UD	late PV (6.3 min)	30	6.864	0.222	3
[76]	7	Road Tunnel	Netherlands	UD	late PV (6.2 min)	30	7.944	0.573	3
[76]	8	Road Tunnel	Netherlands	UD	late PV (5.85 min)	36	6.486	0.215	3

* Data converted from graphical to digital form

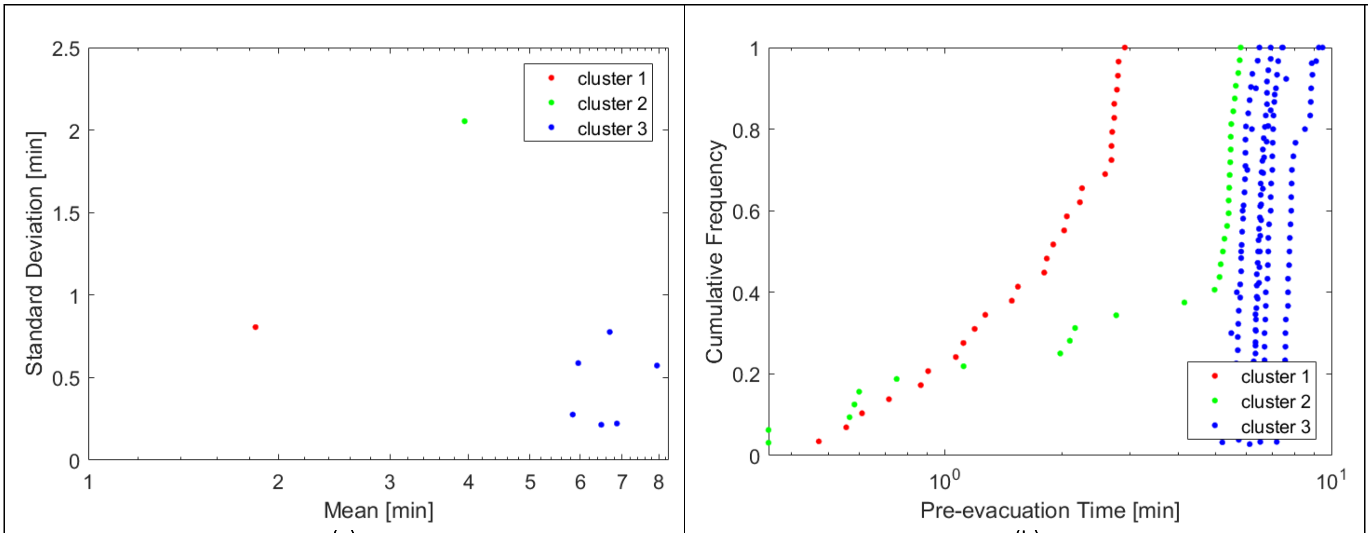

Figure 20 – Pre-evacuation data for road tunnel occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data point of the case studies.

Table 15 – Estimate parameters of the pre-evacuation distributions for the road tunnel clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma*	3.201	35.874	3.201	0.598	29	0.950
	Lognormal*	163.758	0.591	18.410	0.591		0.938
	Loglogistic*	4.620	0.356	0.525	0.356		0.938
	Weibull*	126.413	2.039	2.107	2.039		0.959
2	Gamma*	1.465	185.101	1.465	3.085	32	0.771
	Lognormal*	55.511	0.961	12.732	0.961		0.751
	Loglogistic*	5.333	0.569	1.238	0.569		0.756
	Weibull*	286.860	1.385	4.781	1.385		0.776
3	Gamma**	0.846	90.500	0.846	1.508	161	0.390
	Lognormal**	62.977	1.236	-3.852	1.236		0.399
	Loglogistic**	3.855	0.767	-0.239	0.767		0.397
	Weibull**	74.155	0.890	1.236	0.890		0.391

* it considers the time required to stop the car;

** the reference time (i.e., t=0) is the time when the alarm goes off (the vehicles are already stopped)

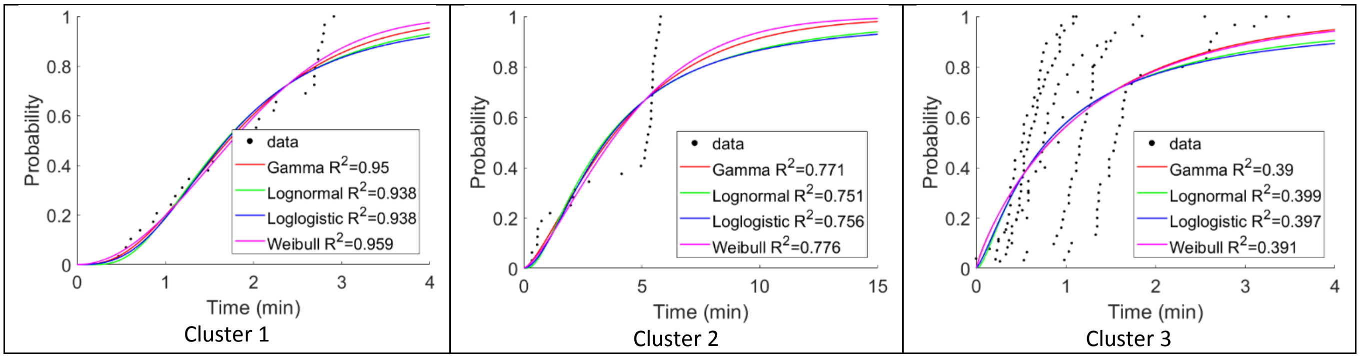

Figure 21 – Pre-evacuation distributions for the road tunnel clusters

The results in Table 15 and Figure 21 show reasonable agreement between the existing data and the proposed distributions for Cluster 1. Cluster 2 has a very specific pattern that cannot be represented with any bi-parametric distributions. In this case, a multiple parameters distribution could have been used or the data could have been truncated into two groups to estimate two different set of distributions. This analysis was not pursued here since, to the best of our knowledge, the most popular and widely used evacuation models are not designed to accommodate such. Hence, it has not been possible to provide usable distributions for Cluster 2. Finally, Cluster 3 has a low value of R^2 as there is a great dispersion of the data belonging to those clusters. However, given the regression charts in Figure 21 the proposed distributions seem to provide a good representation of the trends shown by the observed data.

4.8 Miscellaneous Occupancies

There are 8 case studies belonging to this occupancy group. These evacuations took place in several types of evacuation environments, and in turn, the case studies are divided accordingly to the type of environment without the use of cluster analysis: Ferry (Clusters 1), Cruise Ship (Clusters 2), Hospital Outpatient (Cluster 3), Nuclear Power Plant (Cluster 4), mixed-use buildings including libraries, offices and computer spaces (Cluster 5) and Metro Station (Cluster 6), as illustrated in Table 16.

Figure 22.a illustrates the location of the 8 case studies on the mean vs. standard definition plane while the data points of those case studies are displayed in Figure 22.b.

Table 16– Pre-evacuation data for miscellaneous occupancy

Ref.	Case Study	Environment	Country	Nature	Alarm	Sample	Mean [min]	SD [min]	Cluster
[77]	1	Ferry*	-	UD	AL	553	0.650	0.868	1
[77]	2	Ferry*	-	UD	AL	470	0.861	1.106	1
[77]	3	Cruise Ship*	-	UD	AL	1228	3.131	3.339	2
[72]	4	Hospital Outpatient	UK	UD	AL	33	1.066	0.704	3
[78]	5	Nuclear Power Plant	Sweden	UD	AL	16	1.617	0.926	4
[44]	6	University Library/Office/ Computer Space	UK	UD	AL	153	1.689	0.966	5
[44]	7	University Library/Office/ Computer Space	UK	UD	PV	15	0.870	0.314	5
[44]	8	University Library/Office/ Computer Space	UK	UD	T3	10	0.262	0.158	5
[79]	9	Metro Station	China	UD	-	182	0.453	0.306	6

* Data converted from graphical to digital form

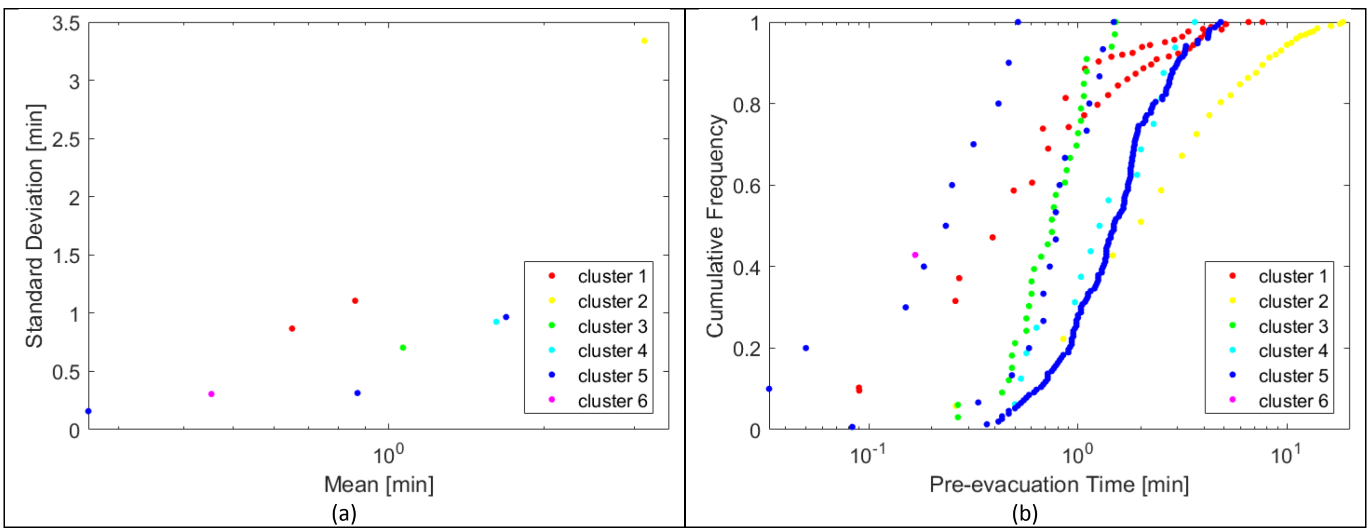

Figure 22 – Pre-evacuation data for miscellaneous occupancy divided by clusters: (a) means and standard deviations of the case studies; (b) data point of the case studies.

Four pre-evacuation distributions were estimated for each cluster. The estimated parameters and the R^2 values are displayed in Table 17. Those distributions and the related data points are illustrated in Figure 23.

Table 17 – Estimate parameters of the pre-evacuation distributions for the miscellaneous clusters

Cluster	Distribution	Distribution in seconds		Distribution in minutes		Data points	R^2
		a	b	a	b		
1	Gamma	0.860	45.026	0.860	0.750	47	0.963
	Lognormal	95.661	1.142	-28.454	1.142		0.980
	Loglogistic	3.167	0.659	-0.927	0.659		0.982
	Weibull	37.179	0.881	0.620	0.881		0.966
2	Gamma	0.891	205.509	0.891	3.425	28	0.994
	Lognormal	332.809	1.078	45.411	1.078		0.998
	Loglogistic	4.748	0.628	0.654	0.628		0.995
	Weibull	177.769	0.922	2.963	0.922		0.995
3	Gamma	5.595	8.460	5.595	0.141	33	0.988
	Lognormal	448.327	0.431	-36.600	0.431		0.987
	Loglogistic	3.785	0.261	-0.309	0.261		0.986
	Weibull	52.013	2.639	0.867	2.639		0.987
4	Gamma	2.055	46.065	2.055	0.768	16	0.979
	Lognormal	138.477	0.744	7.873	0.744		0.973
	Loglogistic	4.344	0.448	0.250	0.448		0.971
	Weibull	103.167	1.537	1.719	1.537		0.980
5	Gamma	1.721	58.304	1.721	0.972	178	0.671
	Lognormal	137.519	0.794	9.094	0.794		0.650
	Loglogistic	4.388	0.481	0.293	0.481		0.656
	Weibull	106.771	1.427	1.780	1.427		0.676
6	Gamma	1.067	16.402	1.067	0.273	4	0.998
	Lognormal	32.483	0.941	-20.980	0.941		0.995
	Loglogistic	2.479	0.558	-1.616	0.558		0.994
	Weibull	17.739	1.044	0.296	1.044		0.998

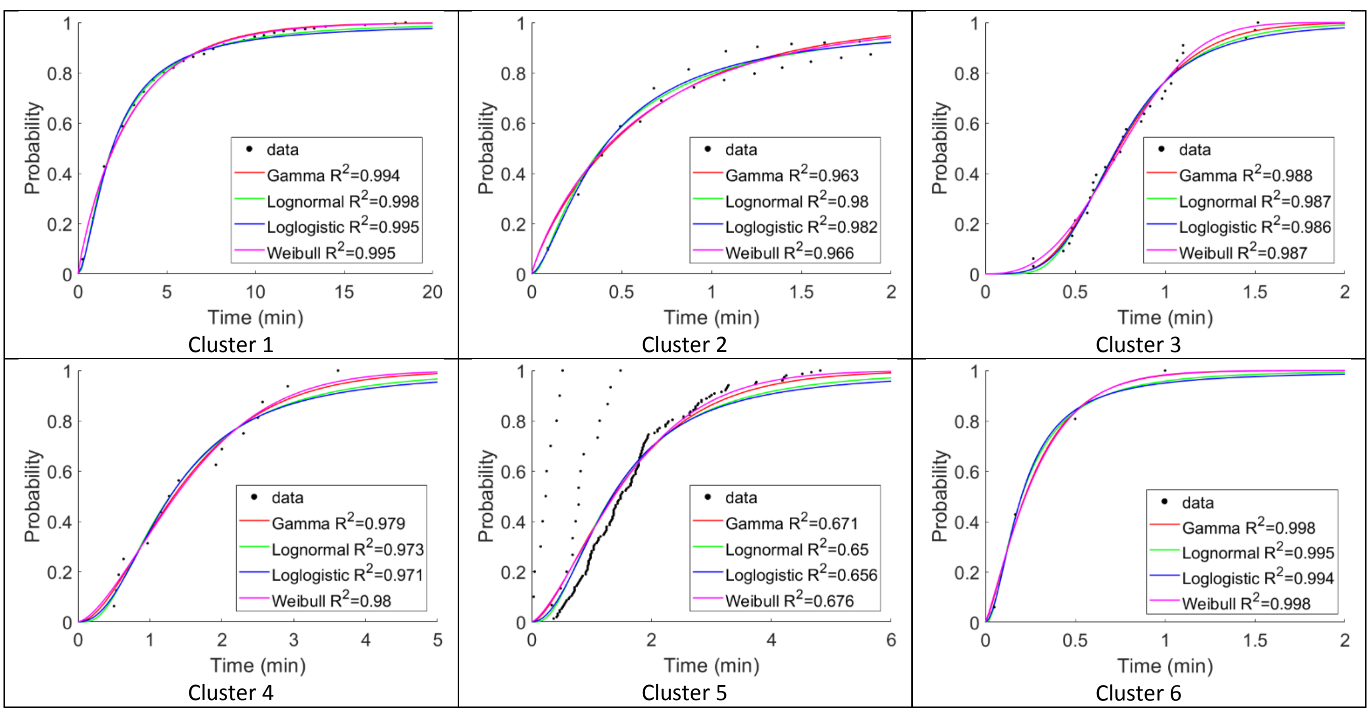

Figure 23 – Pre-evacuation distributions for the miscellaneous clusters

The results In Table 17 and Figure 23 show reasonable agreement between the existing data and the proposed distributions. It is worth highlighting that Cluster 6 refers to the evacuation that took place in a metro station in China. For this evacuation, it was possible to find only four data points from the literature. This small number of data points explains the high value of R^2 in Table 11. Regardless of the limitations of this case study, the proposed distribution is capable of representing the pre-evacuation event as illustrated in Figure 23.

4.9 Result Summary

The summary of the pre-evacuation distributions estimated in this work is provided in Table 18. Those distributions are divided in eight occupancy classes. Users can select among those propose pre-evacuation distributions depending on fire scenarios to simulate. The description column of Table 18 provides a brief description where the data are from. However, we recommend the readers to check for more information regarding those data and the factors affecting those data using the references provided in Sections 4.1-4.8. Moreover, the reader should refer to the instructions in Section 4 on how to select between clusters and how to select the best distribution within the same cluster.

Table 18 – Summary of the pre-evacuation distributions

<table>
<tr><th rowspan="2">Occupancy</th><th rowspan="2">Cluster</th><th rowspan="2">Description</th><th></th><th colspan="2">Distribution in seconds</th><th colspan="2">Distribution in minutes</th><th rowspan="2">R^2</th></tr>
<tr><th>Distribution</th><th>a</th><th>b</th><th>a</th><th>b</th></tr>
<tr><td rowspan="8">Business</td><td rowspan="4">1</td><td rowspan="4">Building: Office and Mixed Office
Country: US, Canada, Finland, UK Denmark, Australia
Nature: UD, AD, P-UD, FI
Alarm: AL, PV and T3
Floors: 4-14</td><td>Gamma</td><td>1.291</td><td>103.901</td><td>1.291</td><td>1.732</td><td>0.564</td></tr>
<tr><td>Lognormal</td><td>381.651</td><td>0.967</td><td>40.919</td><td>0.967</td><td>0.548</td></tr>
<tr><td>Loglogistic</td><td>4.592</td><td>0.587</td><td>0.498</td><td>0.587</td><td>0.548</td></tr>
<tr><td>Weibull</td><td>139.285</td><td>1.195</td><td>2.321</td><td>1.195</td><td>0.566</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">Building: World Trade Center
Country: US
Nature: FI
Alarm: none
Floors: 110</td><td>Gamma</td><td>0.557</td><td>1419.096</td><td>0.557</td><td>23.651</td><td>0.942</td></tr>
<tr><td>Lognormal</td><td>36.131</td><td>1.613</td><td>11.104</td><td>1.613</td><td>0.949</td></tr>
<tr><td>Loglogistic</td><td>5.905</td><td>0.958</td><td>1.811</td><td>0.958</td><td>0.950</td></tr>
<tr><td>Weibull</td><td>672.010</td><td>0.664</td><td>11.200</td><td>0.664</td><td>0.944</td></tr>
<tr><td rowspan="8">Residential</td><td rowspan="4">1</td><td rowspan="4">Building: Apartment, Univ. Residence, Residential Care
Country: US, Canada, UK
Nature: UD
Alarm: AL and T3 (good alarm performance)
Floor: 3-14</td><td>Gamma</td><td>0.650</td><td>178.024</td><td>0.650</td><td>2.967</td><td>0.601</td></tr>
<tr><td>Lognormal</td><td>54.879</td><td>1.432</td><td>-0.119</td><td>1.432</td><td>0.589</td></tr>
<tr><td>Loglogistic</td><td>4.087</td><td>0.873</td><td>-0.007</td><td>0.873</td><td>0.586</td></tr>
<tr><td>Weibull</td><td>102.475</td><td>0.767</td><td>1.708</td><td>0.767</td><td>0.599</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">Building: Apartment
Country: Canada, Australia
Nature: FI, UD
Alarm: AL (early morning and/or poor performance)
Floors: 3-30</td><td>Gamma</td><td>0.911</td><td>812.708</td><td>0.911</td><td>13.545</td><td>0.820</td></tr>
<tr><td>Lognormal</td><td>98.986</td><td>1.268</td><td>32.821</td><td>1.268</td><td>0.785</td></tr>
<tr><td>Loglogistic</td><td>6.143</td><td>0.763</td><td>2.049</td><td>0.763</td><td>0.784</td></tr>
<tr><td>Weibull</td><td>724.617</td><td>0.978</td><td>12.077</td><td>0.978</td><td>0.819</td></tr>
<tr><td rowspan="8">Mercantile</td><td rowspan="4">1</td><td rowspan="4">Building: Marks & Spencer, IKEA stores
Country: UK, Sweden
Nature: UD
Alarm: AL and VA
Floors: 1-3</td><td>Gamma</td><td>3.005</td><td>14.564</td><td>3.005</td><td>0.243</td><td>0.901</td></tr>
<tr><td>Lognormal</td><td>62.874</td><td>0.574</td><td>-7.764</td><td>0.574</td><td>0.895</td></tr>
<tr><td>Loglogistic</td><td>3.660</td><td>0.334</td><td>-0.434</td><td>0.334</td><td>0.893</td></tr>
<tr><td>Weibull</td><td>48.453</td><td>1.957</td><td>0.808</td><td>1.957</td><td>0.905</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">Building: Xin Lian Xin store
Country: China
Nature: AD
Alarm: AL
Floors: 1</td><td>Gamma</td><td>2.535</td><td>34.561</td><td>2.535</td><td>0.576</td><td>0.996</td></tr>
<tr><td>Lognormal</td><td>150.491</td><td>0.610</td><td>7.704</td><td>0.610</td><td>0.996</td></tr>
<tr><td>Loglogistic</td><td>4.309</td><td>0.362</td><td>0.215</td><td>0.362</td><td>0.993</td></tr>
<tr><td>Weibull</td><td>96.470</td><td>1.642</td><td>1.608</td><td>1.642</td><td>0.994</td></tr>
</table>

Assembly	1	Building: Theatre, Cinema-Theatre Country: UK, Sweden Nature: UD Alarm: PV, AL Floors: 1	Gamma	10.584	1.664	10.584	0.028	0.961
			Lognormal	417.150	0.308	-186.262	0.308	0.962
			Loglogistic	2.829	0.184	-1.266	0.184	0.960
			Weibull	19.187	3.554	0.320	3.554	0.954
	2	Building: Cinema-Theatre, Cinema Country: Sweden Nature: UD Alarm: AL, PV Floors: 1	Gamma	16.627	1.898	16.627	0.032	0.886
			Lognormal	798.053	0.248	-154.961	0.248	0.888
			Loglogistic	3.428	0.149	-0.666	0.149	0.889
			Weibull	33.790	4.631	0.563	4.631	0.880
	3	Building: Cinema Country: Sweden Nature: UD Alarm: PV Floors: 1	Gamma	12.779	3.138	12.779	0.052	0.992
			Lognormal	428.761	0.281	-51.083	0.281	0.993
			Loglogistic	3.658	0.166	-0.436	0.166	0.992
			Weibull	43.514	4.008	0.725	4.008	0.986
	4	Building: Theatre Country: UK Nature: UD Alarm: PV Floors: 3	Gamma	5.544	9.949	5.544	0.166	0.999
			Lognormal	904.832	0.427	-36.708	0.427	1.000
			Loglogistic	3.935	0.251	-0.159	0.251	0.999
			Weibull	60.959	2.579	1.016	2.579	0.996
	5	Building: Restaurant, Bar Country: UK, China Nature: UD Alarm: AL+PV Floors: 1-2	Gamma	2.861	12.757	2.861	0.213	0.372
			Lognormal	62.497	0.611	-11.308	0.611	0.378
			Loglogistic	3.466	0.376	-0.628	0.376	0.377
			Weibull	39.969	1.780	0.666	1.780	0.367
Hotel	1	Building: Hotel Country: Netherland Nature: UD Alarm: Phone Message Floors: -	Gamma	2.787	29.503	2.787	0.492	0.974
			Lognormal	423.443	0.631	16.852	0.631	0.978
			Loglogistic	4.262	0.379	0.168	0.379	0.978
			Weibull	90.500	1.790	1.508	1.790	0.968
	2	Building: High-rise hotel Country: US Nature: FI Alarm: none Floors: 13	Gamma	0.567	1276.836	0.567	21.281	0.994
			Lognormal	93.846	1.571	27.487	1.571	0.997
			Loglogistic	5.780	0.946	1.686	0.946	0.996
			Weibull	606.399	0.684	10.107	0.684	0.996
	3	Building: High-rise hotel Country: US Nature: FI Alarm: none Floors: 26	Gamma	0.294	14165.384	0.294	236.085	0.875
			Lognormal	13.082	2.560	4.947	2.560	0.784
			Loglogistic	6.589	1.567	2.495	1.567	0.772
			Weibull	2088.400	0.440	34.806	0.440	0.835
Educational	1	Building: Kindergarten Country: Russia Nature: UD (Autumn or spring and winter) Alarm: AL Floors: >1	Gamma	7.859	24.692	7.859	0.412	0.931
			Lognormal	125.572	0.358	27.017	0.358	0.931
			Loglogistic	5.218	0.211	1.124	0.211	0.931
			Weibull	212.216	3.142	3.537	3.142	0.929
	2	Building: Kindergarten, Pre-school, Primary and Secondary school Country: Russia, Czech Rep., Ireland, Spain Nature: UD, SAD Alarm: AL, Verbal Floors: 1-3	Gamma	0.757	30.111	0.757	0.502	0.377
			Lognormal	34.794	1.294	-20.181	1.294	0.380
			Loglogistic	2.591	0.794	-1.504	0.794	0.380
			Weibull	21.378	0.835	0.356	0.836	0.378
	3	Building: Library Country: Poland, Turkey, Czech Rep., UK	Gamma	2.743	28.042	2.743	0.467	0.937
			Lognormal	364.484	0.624	8.701	0.624	0.934

		Nature: UD Alarm: AL, PV, AL+PV, AL+ PV +LM Floors: 2-3	Loglogistic	4.197	0.373	0.103	0.373	0.935
			Weibull	84.428	1.798	1.407	1.798	0.937
	4	Building: Laboratory Country: UK Nature: UD Alarm: PV	Gamma	1.117	30.385	1.118	0.506	0.386
			Lognormal	29.943	1.108	-8.731	1.108	0.372
			Loglogistic	3.177	0.675	-0.917	0.675	0.371
			Weibull	34.413	1.100	0.574	1.100	0.386
	5	Building: Lecture hall Country: China, Italy, Sweden Nature: AD, UD Alarm: AL, PV	Gamma	1.144	24.612	1.144	0.410	0.528
			Lognormal	40.702	1.015	-15.379	1.015	0.504
			Loglogistic	2.977	0.620	-1.117	0.620	0.501
			Weibull	28.908	1.127	0.482	1.127	0.530
Road Tunnel	1	Building: Road Tunnel Country: Sweden Nature: UD Alarm: PV	Gamma*	3.201	35.874	3.201	0.598	0.950
			Lognormal*	163.758	0.591	18.410	0.591	0.938
			Loglogistic*	4.620	0.356	0.525	0.356	0.938
			Weibull*	126.413	2.039	2.107	2.039	0.959
	2	Building: Road Tunnel Country: Netherlands Nature: UD Alarm: late PV	Gamma*	1.465	185.101	1.465	3.085	0.771
			Lognormal*	55.511	0.961	12.732	0.961	0.751
			Loglogistic*	5.333	0.569	1.238	0.569	0.756
			Weibull*	286.860	1.385	4.781	1.385	0.776
	3	Building: Road Tunnel Country: Netherlands Nature: UD Alarm: late PV	Gamma**	0.846	90.500	0.846	1.508	0.390
			Lognormal**	62.977	1.236	-3.852	1.236	0.399
			Loglogistic**	3.855	0.767	-0.239	0.767	0.397
			Weibull**	74.155	0.890	1.236	0.890	0.391
Miscellaneous	1	Building: Ferry Nature: UD Alarm: AL	Gamma	0.860	45.026	0.860	0.750	0.963
			Lognormal	95.661	1.142	-28.454	1.142	0.980
			Loglogistic	3.167	0.659	-0.927	0.659	0.982
			Weibull	37.179	0.881	0.620	0.881	0.966
	2	Building: Cruise Ship Nature: UD Alarm: AL	Gamma	0.891	205.509	0.891	3.425	0.994
			Lognormal	332.809	1.078	45.411	1.078	0.998
			Loglogistic	4.748	0.628	0.654	0.628	0.995
			Weibull	177.769	0.922	2.963	0.922	0.995
	3	Building: Hospital Outpatient Country: UK Nature: UD Alarm: AL	Gamma	5.595	8.460	5.595	0.141	0.988
			Lognormal	448.327	0.431	-36.600	0.431	0.987
			Loglogistic	3.785	0.261	-0.309	0.261	0.986
			Weibull	52.013	2.639	0.867	2.639	0.987
	4	Building: Nuclear Power Plant Country: Sweden Nature: UD Alarm: AL	Gamma	2.055	46.065	2.055	0.768	0.979
			Lognormal	138.477	0.744	7.873	0.744	0.973
			Loglogistic	4.344	0.448	0.250	0.448	0.971
			Weibull	103.167	1.537	1.719	1.537	0.980
	5	Building: University Library/Office/ Country: UK Computer Space Nature: UD Alarm: AL, PV, T3	Gamma	1.721	58.304	1.721	0.972	0.671
			Lognormal	137.519	0.794	9.094	0.794	0.650
			Loglogistic	4.388	0.481	0.293	0.481	0.656
			Weibull	106.771	1.427	1.780	1.427	0.676
	6	Building: Metro Station Country: China Nature: UD Alarm: -	Gamma	1.067	16.402	1.067	0.273	0.998
			Lognormal	32.483	0.941	-20.980	0.941	0.995
			Loglogistic	2.479	0.558	-1.616	0.558	0.994
			Weibull	17.739	1.044	0.296	1.044	0.998

* it considers the time required to stop the car;

** the reference time (i.e., t=0) is the time when the alarm goes off (the vehicles are already stopped)

5. DISCUSSION AND CONCLUSION

Simulating evacuation scenarios requires several modelling inputs. Pre-evacuation time is an important input since it can have significant impact on evacuation results. Pre-evacuation data is typically scarce, partial and presented in a data structure which can be difficult to use as input into evacuation simulations.

With this work, we address such an issue by presenting a pre-evacuation database expanding the one proposed by Gwynne and Boyce [4]. This was done by including 60 additional case studies. From the best of our knowledge, this expanded database, which includes 9 fire incidents and 103 evacuation drills, is the largest available in the fire safety field. Those fire incidents and evacuation drills have been subdivided using the occupancy classification proposed by Gwynne and Boyce in the SFPE Handbook [4], as well as newly added classifications where appropriate. Differently from existing databases, each case study included in the expanded database has been interrogated to obtain a representative frequency distribution instead of relying simply on summary statistics (see Criterion 4 in Section 3.1). This data is fundamental to estimate pre-evacuation distributions using the approach described in Section 3.2.

The proposed database consists of a summary table for each of the 8 occupancies presented in this paper. Each table provides information regarding occupancy type and country where the evacuations occurred. Moreover, information regarding the nature of the evacuations and the alarm system as well as the sample size [6]and pre-evacuation statistics is provided in those tables. This information provides readers with a context of where the pre-evacuation data is from. Moreover, although the proposed data are from fire accidents and drills, the proposed database could be applied to other egress events such as exposure to chemical or biological agents, active shooter as pre-evacuation data for those events are not available yet in the literature.

Clustering analyses were used to investigate potential groups of case studies sharing 'similar' pre-evacuation time mean and standard deviation. This analysis was done using all of the data examined (i.e. across all of the occupancy types, see Figure 7). In the remaining part of the paper, cluster analysis was used within each occupancy class to identify sub-occupancy groups. Such analysis choice acknowledges the importance of type of occupancies as *"it is likely to be the first factor that the engineer encounters and is likely to form the base assessment of the scenario represented"* [4]. An attempt can be made to explain/interpret the clustering results obtained for each occupancy class. The difficulty in interpreting such results may vary across occupancy classes given the characteristics of the case studies presented. A list of possible factors that can explain those results is provided in Section 4. A main limitation of this study is related to the uncertainty in the interpretation of those clustering results; indeed, it is not always possible to state with certainty which factors had an impact on pre-evacuation timing in the proposed case studies.

The clusters represent datasets of sufficient similarity that they are statistically notable. We make no claims as to underlying nature of their connectivity, given differences and discrepancies in the datasets examined. However, the clusters may allow engineers to identify a curve to use once they have associated their scenario to a particular cluster (e.g., given occupancy type). The clusters may also be a starting point for researchers to generate research questions (e.g., to investigate what the underlying mechanisms that produce these clusters might be). Along with some guidance on the methodology for producing such clusters in the future. The paper also very clearly demonstrates the importance of data collectors documenting and presenting scenario information when presenting results.

[6] In this paper, the term 'sample size' refers to the number of evacuees whose pre-evacuation times are analysed and published. As such, for some case studies the sample size can be less than the total number of evacuees involved in a drill or an accident (i.e., population).

In contrast to other databases, this work also provides readers with a set of estimate pre-evacuation distributions that can be directly used in evacuation models. For each identified cluster, we estimate four bi-parametric distributions: Gamma, Lognormal, Loglogistic, and Weibull. Providing distributions for each cluster (within each occupancy type) will significantly simplify the work of evacuation model users and fire safety engineers when forced to choose the most appropriate pre-evacuation inputs for particular evacuation scenarios (see Section 4.9). The criterion used to select between pre-evacuation distributions referring to the same cluster is the R^2 parameter. This parameter provides information regarding the goodness of fitting of the proposed distributions given the available data. However, R^2 parameters must not be used as a criterion to select between clusters. In fact, this parameter is strongly affected by the number of data-points (i.e., having small numbers of data point can generate very high values of R^2 see Cluster 2 in Section 4.3). Form this study, it possible to evaluate that the distributions selected in Section 3.2 provide similar fitting results excluding few exceptions, i.e., hotel distributions.

Another novelty of this paper is the used of data from different case studies to estimate a single pre-evacuation distribution. This is done by combining the Least Squares Method and weights accounting for the sample size of each dataset (see Equation 1). However, such a calibration solution raises the issue of combining datasets having different uncertainties. In this work, we combined data from several studies which used different methodologies to collect pre-evacuation time, such as closed-circuit television videos, questionnaires and interviews. This produced various levels of measurement uncertainty that cannot be accounted for in the calibration procedure used in this work. Each of those data collection approaches have advantages and disadvantages related to the study of pre-evacuation behaviors. As such, it is not always possible to identify the optimal measurement procedure to collect those data. It is also worth noting that data collected with the same methodology, such as via the use of closed-circuit television cameras, have different levels of uncertainty due to differences in coding procedures among various observers.

Acknowledgements
Steve Gwynne would like to recognize Josanne Lortie and Dr. Ahmed Kashef for their efforts during the review process. Erica Kuligowski and Ruggiero Lovreglio would like to recognize Paul Reneke and Nelson Bryner from NIST for their insights during the review process. Finally, the authors want to thank the reviewers for enhancing the quality of this paper.

Reference

[1] S.M. Gwynne, E.R. Galea, M. Owen, P.J. Lawrence, L. Filippidis, A review of the methodologies used in evacuation modelling, Fire Mater. 23 (1999) 383–388. doi:10.1002/(SICI)1099-1018(199911/12)23:6<383::AID-FAM715>3.0.CO;2-2.

[2] E.D. Kuligowski, R.D. Peacock, B.L. Hoskins, A Review of Building Evacuation Models, 2nd Edition-Technical Note (NIST TN) - 1680, Gaithersburg, 2010.

[3] E.D. Kuligowski, Human behavior in fire, in: SFPE Handb. Fire Prot. Eng., Springer, 2016: pp. 2070–2114.

[4] S.M. V. Gwynne, K.E. Boyce, Engineering Data, in: SFPE Handb. Fire Prot. Eng., Springer New York, New York, NY, 2016: pp. 2429–2551. doi:10.1007/978-1-4939-2565-0_64.

[5] British Standard Institute, PD 7974-6:2004 The application of fire safety engineering principles to fire safety design of buildings: Part 6: Human factors: Life safety strategies - Occupant evacuation, behaviour and condition, 2004.

[6] S.M.V. Gwynne, K.E. Boyce, Representing egress data for engineers, in: Interflam 2016, 2016.

[7] D. Purser, M. Bensilum, Quantification of behaviour for engineering design standards and escape time calculations, Saf. Sci. 38 (2001) 157–182. doi:10.1016/S0925-7535(00)00066-7.

[8] M. Kobes, I. Helsloot, B. de Vries, J.G. Post, Building safety and human behaviour in fire: A literature review, Fire Saf. J. 45 (2010) 1–11. doi:10.1016/j.firesaf.2009.08.005.

[9] N.C. McConnell, K.E. Boyce, J. Shields, E.R. Galea, R.C. Day, L.M. Hulse, The UK 9/11 evacuation study: Analysis of survivors' recognition and response phase in WTC1, Fire Saf. J. 45 (2010) 21–34. doi:10.1016/j.firesaf.2009.09.001.

[10] D. Canter, J. Breaux, Jonathan Sime., Domestic, Multiple Occupancy and Hospital Fires, in: D. Canter (Ed.), Fires Hum. Behav., John Wiley & Sons, New York, 1980: pp. 117–136.

[11] J.. Bryan, Behavioral Response to Fire and Smoke, in: J.D. DiNenno, E. Al. (Eds.), SFPE Handb. Fire Prot. Eng. 3th Ed., National Fire Protection Association, 2002: p. 3.315-3.341.

[12] E.D. Kuligowski, R.D. Peacock, J.D. Averill, Modeling the Evacuation of the World Trade Center Towers on September 11, 2001, Fire Technol. 49 (2011) 65–81. doi:10.1007/s10694-011-0240-y.

[13] E. Kuligowski, Predicting Human Behavior During Fires, Fire Technol. 49 (2013) 101–120. doi:10.1007/s10694-011-0245-6.

[14] S.M. Gwynne, L.M. Hulse, M.J. Kinsey, Guidance for the Model Developer on Representing Human Behavior in Egress Models, Fire Technol. (2015) 1–26. doi:10.1007/s10694-015-0501-2.

[15] M.K. Lindell, R.W. Perry, The Protective Action Decision Model: Theoretical Modifications and Additional Evidence, Risk Anal. 32 (2012) 616–632. doi:10.1111/j.1539-6924.2011.01647.x.

[16] M.F. Sherman, M. Peyrot, L.A. Magda, R.R.M. Gershon, Modeling pre-evacuation delay by evacuees in World Trade Center Towers 1 and 2 on September 11, 2001: A revisit using regression analysis, Fire Saf. J. 46 (2011) 414–424. doi:10.1016/j.firesaf.2011.07.001.

[17] E.D. Kuligowski, D.S. Mileti, Modeling pre-evacuation delay by occupants in World Trade Center Towers 1 and 2 on September 11, 2001, Fire Saf. J. 44 (2009) 487–496. doi:10.1016/j.firesaf.2008.10.001.

[18] G. Proulx, Movement of People: The Evacuation Timing, in: P.J. Di Nenno, D. Drysdale, C.L. Beyler, W.D. Walton (Eds.), SFPE Handb. Fire Prot. Eng. 3th Ed., National Fire Protection Association, 2002: p. 3.342-3.366.

[19] R. Lovreglio, Modelling Decision-Making in Fire Evacuation based on the Random Utility Theory, Politecnico of Bari, Milan and Turin, 2016.

[20] P.A. Reneke, Evacuation Decision Model, 2013. doi:http://dx.doi.org/10.6028/NIST.IR.7914.

[21] R. Lovreglio, E. Ronchi, D. Nilsson, A model of the decision-making process during pre-evacuation, Fire Saf. J. 78 (2015). doi:10.1016/j.firesaf.2015.07.001.

[22] R. Lovreglio, E. Ronchi, D. Nilsson, An Evacuation Decision Model based on perceived risk, social influence and behavioural uncertainty, Simul. Model. Pract. Theory. 66 (2016) 226–242. doi:10.1016/j.simpat.2016.03.006.

[23] E. Retana, M. Spearpoint, Assessment of the New Zealand verification method pre-travel scenarios using the Evacuation Decision Model, in: 6th Int. Symp. Hum. Behav. Fire, 2015.

[24] R. Lovreglio, D. Borri, E. Ronchi, A. Fonzone, L. Dell'Olio, The need of latent variables for modelling decision-making in evacuation simulations, in: IX Int. Work. Plan. Eval., Bari, 2015. doi:10.13140/RG.2.1.4521.7684.

[25] S.M.V. Gwynne, K.E. Boyce, E.D. Kuligowski, D. Nilsson, A. Robbins, R. Lovreglio, Pros and Cons of Egress Drills, in: Interflam 2016, London, UK, 2016.

[26] S.M.V. Gwynne, E.D. Kuligowski, K.E. Boyce, D. Nilsson, A.P. Robbins, R. Lovreglio, J.R. Thomas, A. Roy-Poirier, Enhancing egress drills: Preparation and assessment of evacuee performance, Fire Mater. (2017). doi:10.1002/fam.2448.

[27] L. Shi, Q. Xie, X. Cheng, L. Chen, Y. Zhou, R. Zhang, Developing a database for emergency evacuation model, Build. Environ. 44 (2009) 1724–1729. doi:10.1016/j.buildenv.2008.11.008.

[28] R.F.. Fahy, G. Proulx, Toward creating a database on delay times to start evacuation and walking speeds for use in evacuation modeling, in: Second Int. Symp. Hum. Behav. Fire, 2001.

[29] M. Forssberg, J. Kjellström, Förberedelsetidens variation vid utrymning - LU Report 5543, 2017.

[30] W.H. Greene, Econometric Analysis - 7th Edition, Pearson, 2011.

[31] MathWorks, Weighted Nonlinear Regression, Https://Www.Mathworks.Com/Help/Stats/Examples/Weighted-Nonlinear-Regression.Html. (2017).

[32] H.J. Motulsky, L.A. Ransnas, Fitting curves to data using nonlinear regression: a practical and nonmathematical review., FASEB J. 1 (1987) 365–74. doi:10.1096/FASEBJ.1.5.3315805.

[33] T. Korhonen, Fire Dynamics Simulator with Evacuation: FDS+Evac - VTT Technical Reference and User's Guide, 2017.

[34] Thunderhead Engineering, Pathfinder 2017 - Technical Reference, 2017.

[35] J. Parke, S. Gwynne, E.R. Galea, P. Lawrence, Validating the buildingEXODUS evacuation model using data from an unannounced trial evacuation., in: Pedestr. Evacuation Dyn. 2003, 2003: pp. 295–306.

[36] A.C. Rencher, Methods of multivariate analysis, John Wiley & Sons, 2003.

[37] T.M. Kodinariya, P.R. Makwana., Review on determining number of Cluster in K-Means Clustering, Int. J. Adv. Res. Comput. Sci. Manag. Stud. 1 (2013) 90–95.

[38] J. Bryan, An examination and analysis of the dynamics of the human behavior in the Westchase Hilton Hotel fire, Houston, Texas on March 6, 1982 - NFPA Report, 1983.

[39] J.L. Bryan, A review of the examination and analysis of the dynamics of human behavior in the fire at the MGM Grand Hotel, Clark County, Nevada as determined from a selected questionnaire population, Fire Saf. J. 5 (1983) 233–240. doi:10.1016/0379-7112(83)90021-8.

[40] D.A. Purser, Comparisons of Evacuation Efficiency and Pre-travel Activity Times in Response to a Sounder and Two Different Voice Alarm Messages, in: W.W.F. Klingsch, C. Rogsch, A. Schadschneider, M. Schreckenberg (Eds.), Pedestr. Evacuation Dyn. 2008, Springer-Verlag, 2010: pp. 121–134.

[41] D. Nilsson, H. Frantzich, Design of Voice Alarms—the Benefit of Mentioning Fire and the Use of a Synthetic Voice, in: W.W.F. Klingsch, C. Rogsch, A. Schadschneider, M. Schreckenberg (Eds.), Pedestr. Evacuation Dyn. 2008, Springer, 2010: p. 10.

[42] P. Brennan, Timing Human Response In Real Fires, Fire Saf. Sci. 5 (1997) 807–818.

[43] G.. Proulx, J.. Pineau, J.C.. Latour, L. Stewart, Study of the Occupants' Behaviour During the 2 Forest Laneway Fire in North York, Ontario January 6, 1995 - NRCC Report 705, 1995.

[44] S.M.V. Gwynne, D.L. Boswell, G. Proulx, Understanding the Effectiveness of Notification Technologies in Assisting Vulnerable Populations, J. Fire Prot. Eng. 19 (2009) 31–49. doi:10.1177/1042391508095094.

[45] S.M.V. Gwynne, Optimizing fire alarm notification for high risk groups research project. Notification effectiveness for large groups., 2007.

[46] V. V. Kholshchevnikov, D.A. Samoshin, A.P. Parfyonenko, I.P. Belosokhov, Study of children evacuation from pre-school education institutions, Fire Mater. 36 (2012) 349–366. doi:10.1002/fam.2152.

[47] T.J. Shields, B. Smyth, K.E. Boyce, G.W.H. Silcock, Evacuation behaviours of occupants with learning difficulties in residential homes, Disabil. Rehabil. 21 (1999) 39–48. doi:10.1080/096382899298089.

[48] G. Proulx, N. Bénichou, Evacuation Movement in Photoluminescent Stairwells, in: W.W.F. Klingsch, C. Rogsch, A. Schadschneider, M. Schreckenberg (Eds.), Pedestr. Evacuation Dyn. 2008, Springer-Verlag, 2010: pp. 25–42.

[49] G.. Proulx, A.. Kaufman, J. Pineau, Evacuation Time and Movement in Office Buildings - NRCC Technical Report, 1996.

[50] S. Hostikka, T. Paloposki, T. Rinne, J. Saari, T. Korhonen, Evacuation experiments in offices and public buildings. VTT Technical Research Centre of Finland, Espoo, Finland., 2007.

[51] S.B. Sharma, V. Tabak, D. Brocklehurst, A. Sagun, D. Bouchlaghem, A comprehensive modern approach to developing evacuation data capture/analysis and simulation tools for real world fire engineering, in: Fourth Int. Symp. Hum. Behav. Fire, 2009: pp. 195–206.

[52] B. Christoffersen, C. Söderlind., Comparison of two egress models and a full-scale experiment., in: Fourth Int. Symp. Hum. Behav. Fire, 2009: pp. 573–578.

[53] G.. Proulx, J.C.. Latour, J.W. MacLaurin, Housing Evacuation of Mixed Abilities Occupants - NRCC Report No. 661, 1994.

[54] G. Proulx, J.C. Latour, J.W. MacLaurin, J. Pineau, L.E. Hoffman, C. Laroche, Housing evacuation of mixed abilities occupants in highrise buildings - NRCC Report 706, 1995.

[55] T.J. Shields, K.E. Boyce, G.W.H. Silcock, Towards the characterization of large retail stores, Fire Mater. 23 (1999) 325–331. doi:10.1002/(SICI)1099-1018(199911/12)23:6<325::AID-FAM706>3.0.CO;2-0.

[56] T.. Shields, K.. Boyce, A study of evacuation from large retail stores, Fire Saf. J. 35 (2000) 25–49. doi:10.1016/S0379-7112(00)00013-8.

[57] H. Frantzich, Occupant behaviour and response time - Results from evacuation experiments, in: Second Int. Symp. Hum. Behav. Fire, 2001: pp. 159–166.

[58] C. Xudong, Z. Heping, X. Qiyuan, Z. Yong, Z. Hongjiang, Z. Chenjie, Study of announced evacuation drill from a retail store, Build. Environ. 44 (2009) 864–870. doi:10.1016/j.buildenv.2008.06.012.

[59] E.R. Galea, S.J. Deere, C.G. Hopkin, H. Xie, Evacuation response behaviour of occupants in a large theatre during a live performance, Fire Mater. 41 (2017) 467–492. doi:10.1002/fam.2424.

[60] D. Nilsson, A. Johansson, Social influence during the initial phase of a fire evacuation—Analysis of evacuation experiments in a cinema theatre, Fire Saf. J. 44 (2009) 71–79.

[61] J. Holmström, E. Sävmark, Talat utrymningsmeddelande – optimerad utformning utifrån fullskaleförsök - LU Report 5405, 2013.

[62] B. Zhang, T. Chen, X. Ji, A Model of Pre-evacuation Time: A Case Study of A bar Fire, in: Pedestr. Evacuation Dyn., 2017: pp. 534–537.

[63] M. Kobes, I. Helsloot, B. de Vries, J. Post, Exit choice, (pre-)movement time and (pre-)evacuation behaviour in hotel fire evacuation — Behavioural analysis and validation of the use of serious gaming in experimental research, Procedia Eng. 3 (2010) 37–51. doi:10.1016/j.proeng.2010.07.006.

[64] V. V. Kholshchevnikov, D.A. Samoshin, A.P. Parfyonenko, I.P. Belosokhov, Study of children evacuation from pre-school education institutions, Fire Mater. 36 (2012) 349–366. doi:10.1002/fam.2152.

[65] H. Najmanová, E. Ronchi, An Experimental Data-Set on Pre-school Children Evacuation, Fire Technol. 53 (2017) 1509–1533. doi:10.1007/s10694-016-0643-x.

[66] G.N. Hamilton, P.F. Lennon, J. O'Raw, Human behaviour during evacuation of primary schools: Investigations on pre-evacuation times, movement on stairways and movement on the horizontal plane, Fire Saf. J. 91 (2017) 937–946. doi:10.1016/J.FIRESAF.2017.04.016.

[67] A. Cuesta, S.M.V. Gwynne, The collection and compilation of school evacuation data for model use, Saf. Sci. 84 (2016) 24–36. doi:10.1016/J.SSCI.2015.11.003.

[68] E. Galea, G. Sharp, M. Sauter, S. Deere, L. Filippidis, Investigating the impact of culture on evacuation behaviour: a Polish data-set, in: Fourth Int. Symp. Hum. Behav. Fire, 2012: pp. 62–73.

[69] E.R. Galea, S. M., D. S., L. Filippidis, Investigating the Impact of Culture on Evacuation Behavior - A Turkish Data-Set, Fire Saf. Sci. 10 (2011) 709–722. http://www.iafss.org/publications/fss/10/709 (accessed June 4, 2015).

[70] E.R. Galea, S. Deere, G. Sharp, L. Filippidis, L. Hulse, Investigating the impact of culture on evacuation behaviour, in: 12th Int. Fire Sci. Eng. Conf., 2010: pp. 879–892.

[71] E.R. Galea, G. Sharp, L. Filippidis, S. Deere, M. Sauter, Investigating the impact of culture on evacuation behaviour–A UK Data-Set, in: 13th Int Conf. Interflam, 2014: pp. 893–908.

[72] S. Gwynne, E. Galea, J. Parke, J. Hickson, The Collection Of Pre-Evacuation Times From Evacuation Trials Involving A Hospital Outpatient Area And A University Library Facility, Fire Saf. Sci. 7 (2003) 877–888. doi:10.3801/IAFSS.FSS.7-877.

[73] J. Zhang, W. Song, X. Xu, Experiment and multi-grid modeling of evacuation from a classroom, Phys. A Stat. Mech. Its Appl. 387 (2008) 5901–5909. doi:10.1016/j.physa.2008.06.030.

[74] M. D'Orazio, G. Bernardini, An Experimental Study on the Correlation Between "Attachment to Belongings" "Pre-movement" Time, in: Ulrich Weidmann, U. Kirsch, M. Schreckenberg (Eds.), Pedestr. Evacuation Dyn. 2012, Springer-Verlag, 2014: pp. 167–178.

[75] D. Nilsson, M. Johansson, H. Frantzich, Evacuation experiment in a road tunnel: A study of human behaviour and technical installations, Fire Saf. J. 44 (2009) 458–468. doi:10.1016/j.firesaf.2008.09.009.

[76] A. Norén, J. Winér, Modelling Crowd Evacuation from Road and Train Tunnels-Data and design for faster evacuations - LU Report 5127, 2003.

[77] R. Brown, E. Galea, S. Deere, L. Filippidis, Response time data for large passenger ferries and cruise ships, in: Fourth Int. Symp. Hum. Behav. Fire, 2012: pp. 460–471.

[78] H. Frantzich, D. Nilsson, Evacuation in Complex Environments-An Analysis of Evacuation Conditions at a Tunnel Construction Site., in: Fourth Int. Symp. Hum. Behav. Fire, 2010.

[79] G. Xiao-xia, D. Wei, J. Hong-yu, Study on the Social Psychology and Behaviors in a Subway Evacuation Drill in China, Procedia Eng. 11 (2011) 112–119. doi:10.1016/J.PROENG.2011.04.635.

[80] A.-N. Spiess, N. Neumeyer, An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach, BMC Pharmacol. 10 (2010) 6. doi:10.1186/1471-2210-10-6.